Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood

Magnus, J.R.

Published in:
Journal of Econometrics

Publication date:
1982

Citation for published version (APA):
Magnus, J. R. (1982). Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics*, *19*(2-3), 239-285.

Journal of Econometrics 19 (1982) 239–285. North-Holland Publishing Company

MULTIVARIATE ERROR COMPONENTS ANALYSIS OF LINEAR AND NONLINEAR REGRESSION MODELS BY MAXIMUM LIKELIHOOD

Jan R. MAGNUS*

The London School of Economics, London WC2A 2AE, UK

Received March 1981, final version received December 1981

This paper analyses the estimation by maximum likelihood of multivariate error component models, linear and nonlinear, under various assumptions on the errors. Special attention is given to testing and imposing positivity constraints, and equality and inequality restrictions. Also, the appropriate asymptotic covariance matrices are derived, enabling us, inter alia, to perform Wald tests of various relevant hypotheses.

1. Introduction and summary

Suppose we wish to analyze a combined time-series cross-section of individual firm production. In particular, let there be observations on q firms over T years. For firm i in year t, the demands for the p factors of production are explained by the prices of these factors in year t, the output of firm i in year t, and possibly other variables depending on the form of the ith production function. The following set of nonlinear regression equations describes this situation.

$$\textit{Model I:} \qquad y_{it}=f_i(X_{it},\beta_i)+u_{it}, \quad i=1,\ldots,q, \quad t=1,\ldots,T,$$

where y_{it}, f_i, and u_{it} are $(p,1)$ vectors, the inputs X_{it} are (p,l_i) matrices, and the unknown parameter vectors β_i are $(k_i,1)$ vectors. Model I, clearly, is quite general, and arises in many applications. More specific is the linear model

$$\textit{Model II:} \qquad y_{it}=X_t\beta_i+u_{it}, \quad i=1,\ldots,q, \quad t=1,\ldots,T,$$

where X_t is the same for each firm. This situation arises, for example, when the regressors X_{it} are functions of prices only, and each firm faces the same prices.

*I am grateful to A.D. Woodland for many comments and suggestions. All errors and ambiguities are the author's responsibility.

In this paper we shall study the estimation of Models I and II under various assumptions on the error vectors u_{it}. When combining time-series data with cross-section data, the error components approach suggests itself. In the present case, the errors u_{it} are *vectors*, so that we will have to consider error components *matrices*. Now we can make one of the following three assumptions:

Assumption 1 — One error component matrix. $u_{it} \simeq N_p(0, \Delta)$, where Δ is a positive definite (p, p) matrix.

Assumption 2 — Two error components matrices. $u_{it} = e_t + \varepsilon_{it}$, where e_t and ε_{it} are independently distributed as $e_t \simeq N_p(0, \Gamma)$ and $\varepsilon_{it} \simeq N_p(0, \Delta)$, with Γ positive semidefinite, and Δ positive definite, both of order p.

Assumption 3 — Three error components matrices. $u_{it} = e_t + \zeta_i + \varepsilon_{it}$, where e_t, ζ_i, and ε_{it} are independently distributed as $e_t \simeq N_p(0, \Gamma)$, $\zeta_i \simeq N_p(0, Z)$, and $\varepsilon_{it} \simeq N_p(0, \Delta)$, with Γ and Z positive semidefinite, and Δ positive definite, all of order p.

Assumption 1, of course, is the famous seemingly unrelated regressions case [Zellner (1962)], Assumption 2 has been employed by Chamberlain and Griliches (1975), and Assumption 3 was analyzed by Avery (1977) and Baltagi (1980). Consider now the following two extensions to Assumptions 1 and 2:

Assumption 1a. $u_{it} = N_p(0, \lambda_i \Delta)$, where Δ is a positive definite (p, p) matrix, and $\lambda_i > 0$ $(i = 1, \dots, q)$.

Assumption 2a. $u_{it} = e_t + \varepsilon_{it}$, where e_t and ε_{it} are independently distributed as $e_t = N_p(0, \Gamma)$ and $\varepsilon_{it} \simeq N_p(0, \lambda_i \Delta)$, where Γ is positive semidefinite, Δ positive definite, both of order p, and $\lambda_i > 0$ $(i = 1, \dots, q)$.

The error structure of Assumption 1a is intermediate between Zellner-type estimation for all firms combined $[u_{it} \simeq N(0, \Delta)]$ and Zellner-type estimation for each firm separately $[u_{it} \simeq N(0, \Delta_i)]$. A further rationale for Assumptions 1a and 2a is given in section 8. Assumption 2a is employed in an analysis of substitution possibilities between four types of energy in Dutch manufacturing (6 industries, 19 years, 4 cost shares) by Magnus and Woodland (1980).

The purpose of this paper is to discuss the estimation of the parameters in Models I and II under each of the Assumptions 1, 1a, 2, 2a.[1] The method of

[1] Estimation of Models I and II under Assumption 3, although discussed in the original paper [Magnus (1980, section 9)], is not discussed here in an attempt to somewhat shorten an already too lengthy article. The derivations of the appropriate formulae are, however, similar to the ones presented in the present article.

estimation is maximum likelihood (ML) exclusively. While we are well aware of other methods [Maddala and Mount (1973) compare eleven estimators for the two error components model], we still believe that ML is the most powerful estimation technique available to the profession, and, in this case, leads to results which are computationally feasible.

We shall indicate how the restrictions that Δ be positive definite, Γ positive semidefinite, and λ_i $(i=1,\dots,q)$ positive can be incorporated. Also equality and inequality restrictions on the β's are discussed. Particular attention will be given to the imposition and testing of the restriction that the β-vector in two or more industries is the same. The 'asymptotic covariance matrix' for the β_i *and* the error component matrices will be presented, enabling us, inter alia, to perform Wald tests of the hypothesis $\Gamma=0$.

Three limitations of this article are (i) its confinement to ML estimation, (ii) the fact that we assume no lags in the dependent variables, and (iii) its lack of statistical rigor. No attempt is made to establish statistical properties of the ML estimators and therefore no asymptotic justification for their use is provided.

The literature on *two error components* models includes Balestra and Nerlove (1966) and Nerlove (1971a) who allow lagged dependent variables, and Maddala and Mount (1973) who compare eleven estimators, including ML. Wallace and Hussain (1969) considered the use of a *three error components* linear regression model and developed an Aitken estimator of the coefficient vector based on an estimated variance–covariance matrix. Their paper was extended by Amemiya (1971), Maddala (1971), and Nerlove (1971b). Swamy and Arora (1972) studied small sample properties. Later papers on the three error components model include Balestra (1973), Fuller and Battese (1974), and Mundlak (1978).

All the above papers are concerned with the linear *single equation* (i.e. $p=1$) regression model. A linear *multiple equation* regression model with two error component matrices was estimated by Chamberlain and Griliches (1975) using ML techniques. Their error structure is as in our Assumption 2, but their model is different from Model II and, indeed, not even a special case of Model I. The linear multiple equation regression model with three error components matrices was first considered by Avery (1977) who derived a feasible Aitken estimator of the coefficient vector. This estimator is not ML, and, more seriously, asymptotically inefficient. Baltagi (1980) derived an alternative estimator, also not ML, which he showed to be asymptotically efficient.

The plan of this paper is as follows. Section 2 presents two lemmas concerning the determinant and inverse of matrices with a special structure. Sections 3, 4, and 5 discuss the estimation of Model I under Assumption 2 and, as corollaries, Assumption 1. In section 6 we take up the linear Model

II under Assumptions 2 and 1. Wald tests are discussed in section 7. We shall see that for the linear Model II, the Wald tests are very easy and inexpensive to perform. Sections 3–7 thus analyze Models I and II under Assumptions 1 and 2. In section 8 we present the corresponding theorems for the estimation of Models I and II under Assumptions 1a and 2a. Two appendices conclude the paper. Appendix 1 gives some results on matrix differentiation, traces and the 'duplication matrix'. Appendix 2 contains the proofs.

2. Two useful lemmas

In error component analysis, but not only there,[2] one frequently encounters very large matrices of a special structure. The calculation of the inverse, determinant, and (possibly) eigenvalues of such matrices is troublesome and time-consuming, unless we can use their special structure to reduce the order of the matrices concerned. Lemma 2.1 does precisely that and provides the basis for our subsequent analysis.

Lemma 2.1. *Let M_i $(i=1,\dots,s)$ be symmetric idempotent matrices of order n and rank r_i with $\sum_{i=1}^{s} M_i = I_n$. Let further A_i $(i=1,\dots,s)$ be square matrices (possibly complex) of order p. Define the square matrices*[3]

$$\Omega_1=\sum_{i=1}^{s}(M_i\otimes A_i) \quad \text{and} \quad \Omega_2=\sum_{i=1}^{s}(A_i\otimes M_i),$$

both of order np. Then,

(i) $M_i M_j = 0$ *for all* $i \neq j$; $\sum_{i=1}^{n} r_i = n$;

(ii) *the eigenvalues of Ω_1 and Ω_2 are the eigenvalues of $A_1,\dots,A_s$ with multiplicities $r_1,\dots,r_s$;*

(iii) $|\Omega_1| = |\Omega_2| = \prod_{i=1}^{s} |A_i|^{r_i}$;

(iv) *Ω_1 and Ω_2 are non-singular if and only if all A_i $(i=1,\dots,s)$ are non-singular, in which case*

$$\Omega_1^{-1}=\sum_{i=1}^{s}(M_i\otimes A_i^{-1}) \quad \text{and} \quad \Omega_2^{-1}=\sum_{i=1}^{s}(A_i^{-1}\otimes M_i). \quad \square$$

Some special cases of Lemma 2.1 are well-known. For example, if $p=1$, the matrices A_i reduce to scalars α_i, say, and $\Omega_1=\Omega_2=\sum_{i=1}^{s}\alpha_i M_i$. This is the

[2]See Kanemoto (1980) for an application of our Lemma 2.1 in the analysis of macro-dynamic models.

[3]The symbol $\otimes$ denotes the Kronecker product operator.

structure of the disturbance covariance matrix in the two (with $s=2$) and three ($s=4$) error components model. Baltagi (1980) gives the inverse of Ω_1 and Ω_2 in the case that $A_1, \dots, A_s$ are positive definite, but he does not provide the eigenvalues and the determinant.

Let us next prove the following result.

Lemma 2.2. Let L be a non-singular (q,q) matrix, $q \geqq 2$, A and B square matrices of order p, and a and b $(q,1)$ vectors. The (pq,pq) matrix,

$$G = L \otimes B + ab' \otimes A,$$

has determinant

$$|G| = |L|^p |B|^{q-1} |C|,$$

where

$$C = B + \alpha A \quad \text{and} \quad \alpha = b'L^{-1}a,$$

and, if G is non-singular, its inverse is

$$G^{-1} = L^{-1} \otimes B^{-1} - L^{-1}ab'L^{-1} \otimes E,$$

where

$$\begin{aligned} E = C^{-1}AB^{-1} = B^{-1}AC^{-1} = B^{-1}AB^{-1} \qquad & \text{if} \quad \alpha = 0, \\ = \frac{1}{\alpha}(B^{-1} - C^{-1}) \quad & \text{if} \quad \alpha \neq 0. \quad \square \end{aligned}$$

Note: The non-singularity of L is clearly not sufficient for the non-singularity of G (since B or C may be singular). Neither is it necessary. To see this, let

$$A = B, \qquad L = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad a = b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Then L is singular, but

$$G = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \otimes A \quad \text{is non-singular when } A \text{ is.}$$

A special case of Lemma 2.2 is the following corollary, which will prove useful in the sequel.

Corollary 2.1. *Let Λ be a diagonal (q,q) matrix, $q \geqq 2$, with positive diagonal elements, A and B square matrices of order p, and s_q a $(q,1)$ vector of ones, The (pq,pq) matrix,*

$$G = \Lambda s_q s_q' \Lambda \otimes A + \Lambda \otimes B,$$

has determinant

$$|G| = |\Lambda|^p |C| |B|^{q-1},$$

where

$$C = B + \alpha A \quad \textit{and} \quad \alpha = \operatorname{tr} \Lambda,$$

and, if G is non-singular, its inverse is

$$G^{-1} = \frac{1}{\alpha} s_q s_q' \otimes (C^{-1} - B^{-1}) + \Lambda^{-1} \otimes B^{-1}. \quad \square$$

3. Nonlinear regressions with two error component matrices

In this section (as in sections 4 and 5) we consider a set of nonlinear regression equations.

$$y_{it} = f_i(X_{it}, \beta_i) + u_{it}, \qquad i = 1, \ldots, q, \quad t = 1, \ldots, T, \tag{3.1}$$

where y_{it}, f_i, and u_{it} are $(p,1)$ vectors, the inputs X_{it} are (p, l_i) matrices, and the unknown parameter vectors β_i are $(k_i, 1)$ vectors.

As to the errors u_{it}, the following assumption is made.

Assumption 3.1 — Two error components. The error vectors u_{it} decompose as

$$u_{it} = e_t + \varepsilon_{it},$$

where e_t and ε_{it} are independently distributed as

$$e_t \simeq N_p(0, \Gamma) \quad \text{and} \quad \varepsilon_{it} = N_p(0, \Delta),$$

with Γ positive semidefinite and Δ positive definite,[4] both of order p.

[4] Γ may very well be singular (in fact it may be the null matrix), but Δ has to be non-singular if the disturbance covariance matrix is to be non-singular. See formula (3.8).

From Assumption 3.1 we derive that[5]

$$\begin{aligned} Eu_{it}u'_{js} &= \Gamma+\Delta \quad \text{if} \quad i=j, \quad t=s, \\ &= \Gamma \quad \text{if} \quad i\neq j, \quad t=s, \\ &= 0 \quad \text{if} \quad t\neq s. \end{aligned} \tag{3.2}$$

Define the (p,q) matrices and $(pq,1)$ vectors

$$V_t=(u_{1t}u_{2t},\dots,u_{qt}) \quad \text{and} \quad u_t=\operatorname{vec} V_t, \qquad t=1,\dots,T. \tag{3.3}$$

Then,

$$\begin{aligned} EV_tV'_s=\sum_{i=1}^{q} Eu_{it}u'_{is} &= q(\Gamma+\Delta) \quad \text{if} \quad t=s, \\ &= 0 \quad \text{if} \quad t\neq s. \end{aligned} \tag{3.4}$$

Further, the $(pq,1)$ vectors u_t $(t=1,\dots,T)$ are *independently* distributed as $N(0,\Omega)$, with

$$\Omega=\begin{bmatrix} \Gamma+\Delta & \Gamma & \dots & \Gamma \\ \Gamma & \Gamma+\Delta & \dots & \Gamma \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma & \Gamma & \dots & \Gamma+\Delta \end{bmatrix}=s_qs'_q\otimes\Gamma+I_q\otimes\Delta, \tag{3.5}$$

where s_q is a $(q,1)$ vector of ones. Let

$$M=(1/q)\,s_qs'_q \quad \text{and} \quad W=\Delta+q\Gamma, \tag{3.6}$$

then Ω can be written alternatively as

$$\Omega=M\otimes W+(I_q-M)\otimes\Delta. \tag{3.7}$$

Notice that M is an idempotent symmetric (q,q) matrix with rank 1, and that W is a positive definite (p,p) matrix. Hence, it follows from Lemma 2.1 that

$$|\Omega|=|W|\,|\Delta|^{q-1}, \tag{3.8}$$

and

$$\Omega^{-1}=M\otimes W^{-1}+(I-M)\otimes\Delta^{-1}. \tag{3.9}$$

[5]The symbol $'$ denotes the transpose of a vector or matrix.

One further verifies that

$$EV_t M V_t' = W \quad \text{and} \quad \frac{1}{q-1} EV_t(I-M)V_t' = \Delta. \tag{3.10}$$

We may now formulate the following theorem.

Theorem 3.1. The loglikelihood associated with the nonlinear regression model (3.1) under the Assumption 3.1, is $\mathscr{L} \equiv \sum_{t=1}^{T} \mathscr{L}_t$, *with*

$$\mathscr{L}_t = \gamma - \tfrac{1}{2}\log|W| - \tfrac{1}{2}(q-1)\log|\Delta| \\ - \tfrac{1}{2}\operatorname{tr} V_t M V_t' W^{-1} - \tfrac{1}{2}\operatorname{tr} V_t(I-M)V_t'\Delta^{-1}, \tag{3.11}$$

or equivalently,

$$\mathscr{L}_t = \gamma - \tfrac{1}{2}\log|W| - \tfrac{1}{2}(q-1)\log|\Delta| \\ - \tfrac{1}{2}\sum_{i=1}^{q} u_{it}'\Delta^{-1}u_{it} - \frac{1}{2q}\sum_{i,j=1}^{q} u_{it}'(W^{-1}-\Delta^{-1})u_{jt}, \tag{3.12}$$

where

$$\gamma = -\tfrac{1}{2}pq\log 2\pi, \qquad u_{it} = y_{it} - f_i(X_{it}, \beta_i),$$

and V_t *is defined in (3.3) and M and W in (3.6).* □

Several comments are in order. First, it is readily appreciated that, although the complete disturbance covariance matrix of model (3.1) is of order pqT, the loglikehood only contains matrices of order p. Maximum likelihood estimates of the parameter vectors β_i $(i=1,\dots,q)$ and the error component matrices Γ and Δ are obtained by maximizing $\mathscr{L}$ with respect to the β_i $(i=1,\dots,q)$ and the $p(p+1)$ parameters in Γ and Δ (taking into account the symmetry of Γ and Δ). This involves maximization over $\sum_{i=1}^{q} k_i + p(p+1)$ parameters. This number may be prohibitively large, and for that reason we will concentrate $\mathscr{L}$ with respect to Γ and Δ in the next section. Of course, equality restrictions on the β's (like $\beta_i = \beta_j$) are easily incorporated and also reduce the number of parameters.

Secondly, the maximum likelihood estimates of Γ and Δ, though symmetric, will not necessarily be positive (semi) definite. One way to ensure positive (semi) definiteness is to write

$$\Gamma = L_c L_c' \quad \text{and} \quad \Delta = L_d L_d', \tag{3.13}$$

where L_c and L_d are lower triangular (p,p) matrices, and to maximize $\mathscr{L}$ with respect to the $p(p+1)$ parameters in L_c and L_d, rather than Γ and Δ. One may verify that the class of lower triangular matrices L and the class of positive (semi) definite matrices are in 1–1 correspondence, apart from the sign of the diagonal elements of L. Thus, to ensure a unique correspondence we demand that $l_{ii}^{(c)} \geqq 0$ and $l_{ii}^{(d)} > 0$, where $l_{ii}^{(c)}$ and $l_{ii}^{(d)}$ $(i=1,\ldots,p)$ are the diagonal elements of L_c and L_d, respectively. These inequality constraints can be incorporated in $\mathscr{L}$ as follows.[6] Define

$$\alpha_c \equiv \sum_{i=1}^{p} (|l_{ii}^{(c)}| - l_{ii}^{(c)}) \quad \text{and} \quad \alpha_d \equiv \sum_{i=1}^{p} (|l_{ii}^{(d)}| - l_{ii}^{(d)}). \tag{3.14}$$

Then,

$$\begin{aligned} \alpha_c + \alpha_d &= 0 \quad \text{if} \quad l_{ii}^{(c)} \geqq 0, \quad l_{ii}^{(d)} \geqq 0 \quad \text{for} \quad i=1,\ldots,p, \\ &> 0 \quad \text{otherwise.} \end{aligned} \tag{3.15}$$

Let μ be some large positive number (say 1,000). Then, maximization of

$$\mathscr{L}^* \equiv \mathscr{L} - \mu(\alpha_c + \alpha_d) \tag{3.16}$$

will yield the required maximum likelihood estimates.[7]

Thirdly, inequality constraints on the β-parameters can be incorporated quite painlessly. Suppose we know a priori that β_{ik}, say, is constrained by

$$a_1 \leqq \beta_{ik} \leqq a_2. \tag{3.17}$$

Let α_b be defined as

$$\alpha_b \equiv |\beta_{ik} - a_1| + |\beta_{ik} - a_2| - (a_2 - a_1). \tag{3.18}$$

It is clear that $\alpha_b = 0$ for $a_1 \leqq \beta_{ik} \leqq a_2$, and $\alpha_b > 0$ otherwise. Hence, maximization of

$$\mathscr{L}^* \equiv \mathscr{L} - \mu\alpha_b, \tag{3.19}$$

[6]Incorporating these constraints ensures that the L_c and L_d matrices that maximize $\mathscr{L}$ are unique. While this may not be strictly necessary (the resulting Γ and Δ *will* be unique), it may ease the optimization, in particular in cases where Γ and Δ are close to being singular.

[7]Δ must always be nonsingular, but Γ can (and frequently will) be singular. It may, occasionally, be useful to constrain Γ to a singular matrix. This can be done by setting $l_{pp}^{(c)} = 0$ (the last diagonal element L_c). Note that $l_{11}^{(c)} = 0$ implies more than singularity; it implies that the whole first row and column of Γ are zeroes.

where μ again is some large positive number, will yield the maximum likelihood estimates under the constraint (3.17).

Darrough, Pollak and Wales (1980) estimate a model which is similar to the one presented here, except that q (the number of income classes in their case) differs from year to year, i.e., q depends on t. Let q_t be the q in time period t, and define the (q_t, q_t) and (p, p) matrices

$$M_t = (1/q_t)\, s_{q_t} s'_{q_t} \quad \text{and} \quad W_t = \Delta + q_t \Gamma. \tag{3.20}$$

In this case the loglikelihood is $\mathscr{L} = \sum_{t=1}^{T} \mathscr{L}_t$ with

$$\begin{aligned} \mathscr{L}_t &= \gamma_t - \tfrac{1}{2}\log|W_t| - \tfrac{1}{2}(q_t - 1)\log|\Delta| \\ &\quad - \tfrac{1}{2}\operatorname{tr} V_t M_t V_t' W_t^{-1} - \tfrac{1}{2}\operatorname{tr} V_t(I - M_t)V_t'\Delta^{-1} \\ &= \gamma_t - \tfrac{1}{2}\log|W_t| - \tfrac{1}{2}(q_t - 1)\log|\Delta| \\ &\quad - \tfrac{1}{2}\sum_{i=1}^{q_t} u_{it}'\Delta^{-1}u_{it} - (1/2q_t)\sum_{i,j=1}^{q_t} u_{it}'(W_t^{-1} - \Delta^{-1})u_{jt}, \end{aligned} \tag{3.21}$$

where

$$\gamma_t = -\tfrac{1}{2} p q_t \log 2\pi.$$

A final comment: If $\Gamma = 0$, the model (3.1) reduces to the seemingly unrelated regressions model [Zellner (1962)], as is evident from the following corollary.

Corollary 3.1. $(\Gamma = 0)$. *The loglikelihood associated with the nonlinear regression model (3.1) under the assumption that* $u_{it} \simeq N_p(0, \Delta)$, Δ *positive definite, is* $\mathscr{L} = \sum_{t=1}^{T} \mathscr{L}_t$ *with*

$$\begin{aligned} \mathscr{L}_t &= \gamma - \tfrac{1}{2} q \log|\Delta| - \tfrac{1}{2}\operatorname{tr} V_t V_t' \Delta^{-1} \\ &= \gamma - \tfrac{1}{2} q \log|\Delta| - \tfrac{1}{2}\sum_{i=1}^{q} u_{it}'\Delta^{-1}u_{it}. \quad \square \end{aligned} \tag{3.22}$$

4. The loglikelihood concentrated w.r.t. Γ and Δ

We saw in the previous section that — in the absence of equality restrictions on the β's — the number of parameters in $\mathscr{L}$ is $\sum_{i=1}^{q} k_i + p(p+1)$. This number may be quite large. In order to facilitate the estimation, we shall now maximize $\mathscr{L}$ with respect to Γ and Δ only (keeping the β's fixed). This will yield explicit expressions for the maximizing Γ and Δ, and we may therefore concentrate $\mathscr{L}$ with respect to these two matrices. The concentrated $\mathscr{L}$ then contains only $\sum_{i=1}^{q} k_i$ parameters.

This exercise is only possible, it seems, if q (the number of industries, say) does *not* depend on t. In case the loglikelihood is given by (3.21), we are not able to find explicit expressions for the maximizing Γ and Δ.

Rules of differentiation and some related results that are used in the following are presented in appendix 1.

Starting from (3.11), we have

$$\mathscr{L}_t=\gamma-\tfrac{1}{2}\log|W|-\tfrac{1}{2}(q-1)\log|\Delta|$$
$$-\tfrac{1}{2}\operatorname{tr} V_t M V_t' W^{-1}-\tfrac{1}{2}\operatorname{tr} V_t(I-M)V_t'\Delta^{-1}. \tag{4.1}$$

Totally differentiating with respect to Γ and Δ yields

$$\begin{aligned}\mathrm{d}\mathscr{L}_t &= -\tfrac{1}{2}\operatorname{tr} W^{-1}\mathrm{d}W-\tfrac{1}{2}(q-1)\operatorname{tr}\Delta^{-1}\mathrm{d}\Delta\\ &\quad -\tfrac{1}{2}\operatorname{tr} V_tMV_t'\mathrm{d}W^{-1}-\tfrac{1}{2}\operatorname{tr} V_t(I-M)V_t'\mathrm{d}\Delta^{-1}\\ &= -\tfrac{1}{2}\operatorname{tr} W^{-1}\mathrm{d}W-\tfrac{1}{2}(q-1)\operatorname{tr}\Delta^{-1}\mathrm{d}\Delta\\ &\quad +\tfrac{1}{2}\operatorname{tr} V_tMV_t'W^{-1}(\mathrm{d}W)W^{-1}+\tfrac{1}{2}\operatorname{tr} V_t(I-M)V_t'\Delta^{-1}(\mathrm{d}\Delta)\Delta^{-1}.\end{aligned}$$

Hence,

$$\mathrm{d}\mathscr{L}_t=\tfrac{1}{2}\operatorname{tr} W^{-1}[V_tMV_t'-W]W^{-1}\mathrm{d}W$$
$$+\tfrac{1}{2}\operatorname{tr}\Delta^{-1}[V_t(I-M)V_t'-(q-1)\Delta]\Delta^{-1}\,\mathrm{d}\Delta. \tag{4.2}$$

Let

$$\tilde{W}=W^{-1}\left(\sum_{t=1}^{T} V_tMV_t'-TW\right)W^{-1}, \tag{4.3}$$

and

$$\tilde{\Delta}=\Delta^{-1}\left(\sum_{t=1}^{T} V_t(I-M)V_t'-(q-1)T\Delta\right)\Delta^{-1}. \tag{4.4}$$

Then,

$$\mathrm{d}\mathscr{L}=\sum_{t=1}^{T}\mathrm{d}\mathscr{L}_t=\tfrac{1}{2}\operatorname{tr}\tilde{W}\,\mathrm{d}W+\tfrac{1}{2}\operatorname{tr}\tilde{\Delta}\,\mathrm{d}\Delta.$$

Since we wish to maximize $\mathscr{L}$ with respect to Γ and Δ, we substitute $W=\Delta+q\Gamma$, so that

$$\begin{aligned}\mathrm{d}\mathscr{L}&=\tfrac{1}{2}\operatorname{tr}\tilde{W}\mathrm{d}(\Delta+q\Gamma)+\tfrac{1}{2}\operatorname{tr}\tilde{\Delta}\,\mathrm{d}\Delta\\&=q/2\operatorname{tr}\tilde{W}\mathrm{d}\Gamma+\tfrac{1}{2}\operatorname{tr}(\tilde{W}+\tilde{\Delta})\,\mathrm{d}\Delta\\&=q/2(\operatorname{vec}\mathrm{d}\Gamma)'\operatorname{vec}\tilde{W}+\tfrac{1}{2}(\operatorname{vec}\mathrm{d}\Delta)'\operatorname{vec}(\tilde{W}+\tilde{\Delta}).\end{aligned}$$

The symmetry of Γ and Δ is properly taken into account only if we differentiate with respect to the 'essential' elements of Γ and Δ, i.e., with respect to $v(\Gamma)$ and $v(\Delta)$.[8] We therefore write

$$\operatorname{vec}\mathrm{d}\Gamma=D\,\mathrm{d}v(\Gamma)\quad\text{and}\quad\operatorname{vec}\mathrm{d}\Delta=D\,\mathrm{d}v(\Delta),$$

where D is the duplication matrix defined in (A.1). This yields

$$\mathrm{d}\mathscr{L}=q/2(D\,\mathrm{d}v(\Gamma))'\operatorname{vec}\tilde{W}+\tfrac{1}{2}(D\,\mathrm{d}v(\Delta))'\operatorname{vec}(\tilde{W}+\tilde{\Delta}),$$

and hence

$$\mathrm{d}\mathscr{L}=q/2(\mathrm{d}v(\Gamma))'D'\operatorname{vec}\tilde{W}+\tfrac{1}{2}(\mathrm{d}v(\Delta))'D'\operatorname{vec}(\tilde{W}+\tilde{\Delta}).\tag{4.5}$$

Necessary for a maximum is that $\mathrm{d}\mathscr{L}=0$ for all $\mathrm{d}v(\Gamma)\neq 0$ and $\mathrm{d}v(\Delta)\neq 0$. This gives

$$D'\operatorname{vec}\tilde{W}=0\quad\text{and}\quad D'\operatorname{vec}(\tilde{W}+\tilde{\Delta})=0.$$

Because $\tilde{W}$ and $\tilde{\Delta}$ are symmetric, we may write this as

$$D'Dv(\tilde{W})=0\quad\text{and}\quad D'Dv(\tilde{W}+\tilde{\Delta})=0,$$

that is

$$v(\tilde{W})=0\quad\text{and}\quad v(\tilde{W}+\tilde{\Delta})=0,$$

since $D'D$ is non-singular. This implies that

$$\tilde{W}=0\quad\text{and}\quad\tilde{\Delta}=0,$$

and hence, by (4.3) and (4.4),

$$W=\frac{1}{T}\sum_{t=1}^{T}V_tMV_t'\quad\text{and}\quad\Delta=\frac{1}{(q-1)T}\sum_{t=1}^{T}V_t(I-M)V_t'.\tag{4.6}$$

[8] $v(\Gamma)$ contains the elements on and below the diagonal of Γ. See appendix 1 for details.

We can now concentrate the loglikelihood by substituting the expressions for W and Δ from (4.6) into $\mathscr{L}$, using (4.1). This yields

$$\begin{aligned}
\mathscr{L} = \sum_{t=1}^{T} \mathscr{L}_t &= \gamma T - \tfrac{1}{2} T\log\left|\frac{1}{T}\sum_{t=1}^{T} V_t M V_t'\right| \\
&\quad - \tfrac{1}{2}(q-1)T\log\left|\frac{1}{(q-1)T}\sum_{t=1}^{T} V_t(I-M)V_t'\right| - \tfrac{1}{2}pT - \tfrac{1}{2}p(q-1)T \\
&= \tfrac{1}{2}pq\,T\left(\log T + \frac{q-1}{q}\log(q-1) - \log 2\pi - 1\right) \\
&\quad - \tfrac{1}{2}T\left(\log\left|\sum_{t=1}^{T} V_t M V_t'\right| + (q-1)\log\left|\sum_{t=1}^{T} V_t(I-M)V_t'\right|\right).
\end{aligned}$$

Summarizing, we have proved the following theorem.

Theorem 4.1. Consider the model (3.1) under the Assumption 3.1. Maximum likelihood estimates of the parameter vectors $\beta_1,\ldots,\beta_q$ are obtained by minimizing the scalar function

$$\phi(\beta_1,\ldots,\beta_q) = \log\left|\sum_{t=1}^{T} V_t M V_t'\right| + (q-1)\log\left|\sum_{t=1}^{T} V_t(I-M)V_t'\right|, \tag{4.7}$$

with respect to $\beta_1,\ldots,\beta_q$, where

$$V_t = [u_{1t}(\beta_1),\ldots,u_{qt}(\beta_q)], \qquad t=1,\ldots,T,$$

are the (p,q) matrices of disturbances defined in (3.3).

If $\hat\beta_1,\ldots,\hat\beta_q$ are the ML estimates so obtained, and

$$\hat V_t = [u_{1t}(\hat\beta_1),\ldots,u_{qt}(\hat\beta_q)],$$

then the ML estimates of Δ, Γ and W are

$$\hat\Delta = \frac{1}{(q-1)T}\sum_{t=1}^{T} \hat V_t(I-M)\hat V_t', \tag{4.8}$$

$$\hat\Gamma = \frac{1}{q(q-1)T}\sum_{t=1}^{T} \hat V_t(qM-I)\hat V_t', \tag{4.9}$$

$$\hat{W}=\frac{1}{T}\sum_{t=1}^{T}\hat{V}_t M\hat{V}_t', \tag{4.10}$$

and the value of the maximized loglikelihood is

$$\hat{\mathscr{L}}=\tfrac{1}{2}pqT\left[\log T+\frac{q-1}{q}\log(q-1)-\log 2\pi-1\right]$$

$$-\tfrac{1}{2}T\phi(\hat{\beta}_1,\ldots,\hat{\beta}_q).\quad\square \tag{4.11}$$

Theorem 4.1 is a powerful result that simplifies the estimation considerably. There is, however, one problem. While the estimated $\hat{\Delta}$ and $\hat{W}$ are always positive (semi) definite, this is not the case with $\hat{\Gamma}$. Thus, if minimization of (4.7) yields a $\hat{\Gamma}$ which is not positive semidefinite, we have to employ the method of section 3 with $\Gamma=L_c L_c'$ and $\Delta=L_d L_d'$.

When $p=1$, the error structure reduces to the Balestra–Nerlove (1966) structure, and Γ and Δ reduce to scalars. From (4.8) and (4.9) we find

$$\hat{\delta}=\frac{1}{(q-1)T}\left[\sum_{t=1}^{T}\sum_{i=1}^{q}u_{it}^2-\frac{1}{q}\sum_{t=1}^{T}\left(\sum_{i=1}^{q}u_{it}\right)^2\right], \tag{4.12}$$

and

$$\hat{\gamma}=\frac{1}{q(q-1)T}\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{q}u_{it}\right)^2-\sum_{t=1}^{T}\sum_{i=1}^{q}u_{it}^2\right]. \tag{4.13}$$

Balestra and Nerlove do not use γ and δ as their parameters, but

$$\sigma^2\equiv\gamma+\delta\quad\text{and}\quad\rho\equiv\gamma/(\gamma+\delta).$$

It is clear that $\hat{\gamma}$ (and hence $\hat{\rho}$) may be negative. It is not surprising therefore that our more general model possesses this same property.

An approach which seems to suggest itself, is to maximize $\mathscr{L}$ with respect to Δ only. Unfortunately the ML condition in that case does not lead to an explicit expression for Δ.

If $\Gamma=0$ a priori, we have the following well-known corollary.

Corollary 4.1 ($\Gamma=0$). *Consider the model (3.1) under the assumption that $u_{it}\simeq N(0,\Delta)$, Δ positive definite. Maximum likelihood estimates of the parameter vectors $\beta_1,\ldots,\beta_q$ are obtained by minimizing the scalar function*

$$\phi(\beta_1,\ldots,\beta_q)=\log\left|\sum_{t=1}^{T}V_tV_t'\right|,$$

with respect to $\beta_1, \ldots, \beta_q$. *The ML estimate of* Δ *is*

$$\hat{\Delta} = \frac{1}{qT} \sum_{t=1}^{T} V_t V_t',$$

and the value of the maximized loglikelihood is

$$\hat{\mathscr{L}} = \tfrac{1}{2} pqT[\log qT - \log 2\pi - 1] - \tfrac{1}{2} qT\phi(\hat{\beta}_1, \ldots, \hat{\beta}_q). \quad \square$$

5. The information matrix

Let us abbreviate the parameters of our model (3.1) as

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}, \qquad \theta = \begin{pmatrix} v(\Gamma) \\ v(\Delta) \end{pmatrix}, \qquad \eta = \begin{pmatrix} \beta \\ \theta \end{pmatrix}, \tag{5.1}$$

where β, θ, and η are column vectors of order $\sum_{i=1}^{q} k_i$, $p(p+1)$, and $\sum_i k_i + p(p+1)$, respectively.

The precision of the ML estimators $\hat{\beta}$ and $\hat{\theta}$ can be stated in terms of the information matrix defined by

$$\Psi_T \equiv -E(\partial^2 \mathscr{L} / \partial\eta\, \partial\eta').$$

The Cramér–Rao inequality tells us that, under general conditions, Ψ_T^{-1} is a lower bound for the covariance matrix of any unbiased estimator of β and θ. The asymptotic information matrix is defined as

$$\Psi \equiv \lim_{T \to \infty} (1/T)\, \Psi_T,$$

and, since for $T \to \infty$,

$$\sqrt{T}(\hat{\eta} - \eta) \to N(0, \Psi^{-1}),$$

we may, for finite T, use Ψ_T^{-1} as an estimator for $\operatorname{cov}(\hat{\eta})$. Ψ_T^{-1} is usually referred to as 'the asymptotic covariance matrix' of $\hat{\eta}$.

The information matrix Ψ_T is block-diagonal in β and θ [see Heymans and Magnus (1979)], so that the asymptotic covariance matrix for θ can be obtained independently from that of β.

Theorem 5.1 — Asymptotic covariance matrix of $\hat{\beta}$. *Consider the model (3.1) under the Assumption 3.1. Define the* (p, k_i) *matrices*

$$H_{it} = (\partial f_i(X_{it}, \beta_i)/\partial \beta_i)', \qquad i=1,\dots,q, \quad t=1,\dots,T, \tag{5.2}$$

and the (k_i, k_j) *and* (k_i, k_i) *matrices*

$$C_{ij} = \sum_{t=1}^{T} H'_{it}(W^{-1} - \Delta^{-1})H_{jt}, \qquad i,j=1,\dots,q, \tag{5.3}$$

and

$$D_i = \sum_{t=1}^{T} H'_{it}\Delta^{-1}H_{it}, \qquad i=1,\dots,q. \tag{5.4}$$

Then the information matrix for $\beta = (\beta'_1, \dots, \beta'_q)$ *is*

$$\Psi_T^{(\beta)} = \frac{1}{q}\begin{bmatrix} C_{11} & \cdots & C_{1q} \\ \vdots & & \vdots \\ C_{q1} & \cdots & C_{qq} \end{bmatrix} + \begin{bmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_q \end{bmatrix}, \tag{5.5}$$

and its inverse is the 'asymptotic covariance matrix' of $\hat{\beta}$. □

Its counterpart for $\hat{\theta}$ is contained in Theorem 5.2.

Theorem 5.2 — Asymptotic covariance matrix of $\hat{\theta}$. *Consider the model (3.1) under the Assumption 3.1. Let* $P(A)$ *and* $R(A)$ *denote the following matrix functions of a* (p, p) *matrix* A:

$$P(A) = D'(A \otimes A)D, \qquad R(A) = D^{+}(A \otimes A)D^{+\prime}. \tag{5.6}$$

The information matrix for $v(\Gamma)$ *and* $v(\Delta)$ *is the* $(p(p+1), p(p+1))$ *matrix*

$$\Psi_T^{(\theta)} = \tfrac{1}{2}T\begin{bmatrix} q^2P(W^{-1}) & qP(W^{-1}) \\ qP(W^{-1}) & P(W^{-1}) + (q-1)P(\Delta^{-1}) \end{bmatrix}, \tag{5.7}$$

and the 'asymptotic covariance matrix' of the ML estimators $v(\hat{\Gamma})$ *and* $v(\hat{\Delta})$ *is*

$$\Psi_T^{(\theta)-1} = \frac{2}{q^2(q-1)T}\begin{bmatrix} R(\Delta) + (q-1)R(W) & -qR(\Delta) \\ -qR(\Delta) & q^2R(\Delta) \end{bmatrix}. \quad \square \tag{5.8}$$

Given the asymptotic covariance matrix of $v(\hat{\Gamma})$, we can test the null hypothesis $\Gamma=0$, using the Wald statistic

$$(q^2(q-1)T/2)\, v(\Gamma)'[R(\Delta)+(q-1)R(W)]^{-1}v(\Gamma), \tag{5.9}$$

which is asymptotically χ^2 distributed with $\frac{1}{2}p(p+1)$ degrees of freedom.

If $\Gamma=0$ a priori, Theorems 5.1 and 5.2 reduce to:

Corollary 5.1 ($\Gamma=0$). *Consider the model (3.1) under the assumption that $u_{it} \simeq N(0, \Delta)$, Δ positive definite. The information matrices for β and $v(\Delta)$ are*

$$\Psi_T^{(\beta)} = \begin{bmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_q \end{bmatrix} \quad \text{and} \quad \Psi_T^{(\Delta)} = \tfrac{1}{2}qTP(\Delta^{-1}), \tag{5.10}$$

where D_i $(i=1,\ldots,q)$ is defined in (5.4) and $P(\Delta^{-1})$ in (5.6). The 'asymptotic covariance matrix' is

$$\Psi_T^{-1} = \left[\begin{array}{ccc|c} D_1^{-1} & & 0 & 0 \\ & \ddots & & \vdots \\ & & D_q^{-1} & 0 \\ \hline 0 & \ldots & 0 & \dfrac{2}{qT}R(\Delta) \end{array}\right]. \quad \square \tag{5.11}$$

6. A linear case: Concentration w.r.t. β

The linear version of the nonlinear regression model (3.1) is

$$y_{it} = X_{it}\beta_i + u_{it}, \qquad i=1,\ldots,q, \quad t=1,\ldots,T, \tag{6.1}$$

a special case of which is

$$y_{it} = X_t\beta_i + u_{it}, \qquad i=1,\ldots,q, \quad t=1,\ldots,T, \tag{6.2}$$

where y_{it} and u_{it} are $(p,1)$ vectors, the inputs X_t are (p,k) matrices, and the unknown parameter vectors β_i are $(k,1)$ vectors.

It is easy to think of many examples where model (6.2) is appropriate. In Magnus and Woodland (1980), e.g., y_{it} is a vector of p cost shares for industry i in year t. The cost shares are linear in the input prices, and since each industry faces the same prices, this leads to (6.2).

As to the error vectors u_{it}, Assumption 3.1 (two error component matrices) is maintained. Specification (6.2) is attractive because, as we shall see, explicit and simple expressions can be derived for the β_i ($i=1,\dots,q$) that maximize the likelihood. In other words, it enables us to concentrate the loglikelihood with respect to $\beta_1,\dots,\beta_q$.

Let us write (6.2) in stacked form as

$$y_t=(I_q\otimes X_t)\beta+u_t, \qquad t=1,\dots,T, \tag{6.3}$$

where

$$y_t=\begin{bmatrix} y_{1t} \\ \vdots \\ y_{qt} \end{bmatrix}, \qquad \beta=\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_q \end{bmatrix}, \qquad u_t=\begin{bmatrix} u_{1t} \\ \vdots \\ u_{qt} \end{bmatrix}. \tag{6.4}$$

There may be restrictions on the β_i that we want to incorporate in the model (and test, see section 7). Only simple restrictions of the type $\beta_i=\beta_j=\beta_k$ are considered here, and these are formalized as follows.

Definition 6.1. Let I be the set of integers $(1,2,\dots,q)$. Then r mutually exclusive and exhaustive subsets π_h ($h=1,\dots,r$) can be defined as

$$\pi_1=(1,2,\dots,q_1),$$

$$\pi_2=(q_1+1,q_1+2,\dots,q_1+q_2),$$

$$\vdots$$

$$\pi_r=(q_1+q_2+\cdots+q_{r-1}+1,\dots,q_1+q_2+\cdots+q_r),$$

with

$$\sum_{h=1}^{r} q_h=q.$$

Note that π_h contains q_h integers.

Assumption 6.1. For $i=1,2,\dots,q$ and $h=1,2,\dots,r$,

$$\beta_i=\tilde{\beta}_h \quad \text{if and only if} \quad i\in\pi_h.$$

The assumption simply states that, instead of q β-vectors, we only have r. These we denote as $\tilde{\beta}_1,\tilde{\beta}_2,\dots,\tilde{\beta}_r$. For example, suppose $q=8$ and let the restrictions be $\beta_1=\beta_2$ and $\beta_3=\beta_4=\beta_5=\beta_6$. Then $r=4$, and $q_1=2$, $q_2=4$, $q_3=q_4=1$. Further, $\tilde{\beta}_1=\beta_1=\beta_2$, $\tilde{\beta}_2=\beta_3=\beta_4=\beta_5=\beta_6$, $\tilde{\beta}_3=\beta_7$, $\tilde{\beta}_4=\beta_8$.

Let

$$\tilde{\beta}=(\tilde{\beta}'_1,\ldots,\tilde{\beta}'_r)',$$

a $(kr,1)$ vector, then β and $\tilde{\beta}$ are related through

$$\beta=(N\otimes I_k)\tilde{\beta}, \tag{6.5}$$

where N is the blockdiagonal (q,r) matrix

$$N=\begin{bmatrix} s_{q_1} & & & 0 \\ & s_{q_2} & & \\ & & \ddots & \\ 0 & & & s_{q_r} \end{bmatrix}, \tag{6.6}$$

and s_{q_h} is a vector of q_h ones. In the above example N takes the form

$$N=\left[\begin{array}{cccc} 1 & & & \\ 1 & & & \\ \hline & 1 & & \\ & 1 & & \\ & 1 & & \\ & 1 & & \\ \hline & & 1 & \\ \hline & & & 1 \end{array}\right],$$

where all undesignated elements are zero. Combining (6.3) and (6.5) the model now reads

$$y_t=(N\otimes X_t)\tilde{\beta}+u_t, \qquad t=1,\ldots,T. \tag{6.7}$$

Theorem 6.1. Consider the linear regression model (6.2) under the Assumptions 3.1 (on the error vectors) and 6.1 (on equality constraints among the β's). The maximum likelihood equations for the $\tilde{\beta}_h\,(h=1,\ldots,r)$ are given by

$$\tilde{\beta}_h=A^{-1}a+B^{-1}(b_h-b), \qquad h=1,\ldots,r. \tag{6.8}$$

The 'asymptotic covariance matrix' of the ML estimators $\hat{\tilde{\beta}}_h$ $(h=1,\ldots,r)$ is described by

$$\text{as.var}(\hat{\tilde{\beta}}_h)=(1/q)(A^{-1}-B^{-1})+(1/q_h)B^{-1},$$

and

$$\text{as.cov}(\hat{\tilde{\beta}}_i; \hat{\tilde{\beta}}_h) = (1/q)(A^{-1} - B^{-1}), \qquad i \neq h,$$

where

$$A = \sum_{t=1}^{T} X_t' W^{-1} X_t, \qquad B = \sum_{t=1}^{T} X_t' \Delta^{-1} X_t,$$

$$a = \sum_{t=1}^{T} X_t' W^{-1} \bar{y}_t, \qquad b = \sum_{t=1}^{T} X_t' \Delta^{-1} \bar{y}_t, \qquad b_h = \sum_{t=1}^{T} X_t' \Delta^{-1} \bar{y}_{ht},$$

$$\bar{y}_t = (1/q) \sum_{j=1}^{q} y_{jt}, \qquad \bar{y}_{ht} = (1/q_h) \sum_{j \in \pi_h} y_{jt}, \qquad h = 1, \ldots, r. \quad \square$$

Two special cases are of interest. If $r = q$, there are no restrictions on the β_j and $N = I_q$. The ML equations for the β_j ($j = 1, \ldots, q$) are then given by

$$\beta_j = \left(\sum_{t=1}^{T} X_t' W^{-1} X_t \right)^{-1} \sum_{t=1}^{T} X_t' W^{-1} \bar{y}_t + \left(\sum_{t=1}^{T} X_t' \Delta^{-1} X_t \right)^{-1} \sum_{t=1}^{T} X_t' \Delta^{-1} (y_{jt} - \bar{y}_t). \tag{6.9}$$

If $r = 1$, all β_j are constrained to be equal and $N = s_q$. Then,

$$\beta_1 = \beta_2 = \cdots = \beta_q = \left(\sum_{t=1}^{T} X_t' W^{-1} X_t \right)^{-1} \sum_{t=1}^{T} X_t' W^{-1} \bar{y}_t. \tag{6.10}$$

There are at least four ways in which the $rk + p(p+1)$ parameters ($\tilde{\beta}_1, \ldots, \tilde{\beta}_r$, Γ, Δ) of the linear model (6.2) can be estimated. First, maximization of the complete loglikelihood is possible (Theorem 3.1). This involves maximization over all parameters. The loglikelihood, however, is very simple. Also, the constraint that Γ and Δ be positive (semi)definite is easily imposed by letting $\Gamma = L_c L_c'$ and $\Delta = L_d L_d'$, as described in section 3. A second possibility is to maximize $\mathscr{L}$ concentrated with respect to Γ and Δ (see section 4). The number of parameters then reduces to rk, but, as noted before, there is no guarantee that $\hat{\Gamma}$ is positive semidefinite. A new possibility, which did not exist in the nonlinear model, is to concentrate $\mathscr{L}$ with respect to $\tilde{\beta}_1, \ldots, \tilde{\beta}_r$ and then maximize the concentrated likelihood with respect to Γ and Δ. Thus, the concentrated likelihood is given by (3.11) or (3.12) with V defined in terms of the ML $\tilde{\beta}$'s. This involves maximization over $p(p+1)$ parameters only. Moreover, the requirement that Γ and Δ be positive (semi) definite can be taken into account, as before, by letting $\Gamma = L_c L_c'$ and $\Delta = L_d L_d'$. A fourth possibility is to 'zigzag' between the explicit solution of $\tilde{\beta}_h$ $(h=1,\dots,r)$ in terms of Γ and Δ given by (6.8), and the explicit solution of Γ and Δ in terms of the $\tilde{\beta}_h$ [formulae (4.8) and (4.9)].

The 'zigzag' procedure is — in our experience — the cheapest of the four, but obviously does not guarantee a positive semidefinite $\hat{\Gamma}$. For that reason, the third possibility (concentration w.r.t. $\beta_1,\dots,\beta_r$) is our favourite. Indeed, optimization of a nonlinear function over 12 $(p=3)$ or even 20 $(p=4)$ or 30 $(p=5)$ variables is a straightforward computer routine.

If Γ is known to be the null matrix, Theorem 6.1 reduces to:

Corollary 6.1 $(\Gamma=0)$. *Consider the linear regression model (6.2) under the assumption that $u_{it} \simeq N(0,\Delta)$, Δ positive definite, and Assumption 6.1. The maximum likelihood equations for the $\tilde{\beta}_h$ $(h=1,\dots,r)$ are given by*

$$\tilde{\beta}_h = \left(\sum_{t=1}^{T} X_t' \Delta^{-1} X_t \right)^{-1} \sum_{t=1}^{T} X_t' \Delta^{-1} \bar{y}_{ht}, \qquad h=1,\dots,r,$$

where

$$\bar{y}_{ht} = (1/q_h) \sum_{j \in \pi_h} y_{jt}, \qquad h=1,\dots,r.$$

Further,

$$\text{as.var}\,(\hat{\tilde{\beta}}_h) = (1/q_h) \left(\sum_{t=1}^{T} X_t' \Delta^{-1} X_t \right)^{-1}, \qquad h=1,\dots,r,$$

and the as.cov $(\hat{\tilde{\beta}}_i; \hat{\tilde{\beta}}_h) = 0$ *for* $i \neq h = 1,\dots,r$. □

7. Wald tests

The linear and nonlinear regression models with two error component matrices, as discussed in the previous sections, suggest two hypotheses of a different nature. The first hypothesis is

$$H_0 : \Gamma = 0,$$

and can be tested using the Wald statistic (5.9), or the appropriate likelihood ratio statistic. In both cases the statistic is asymptotically χ^2 distributed with $\frac{1}{2}p(p+1)$ degrees of freedom.

The second hypothesis (or rather set of hypotheses) concerns equality restrictions on the parameter vectors $\beta_1, \beta_2, \dots, \beta_q$. In the nonlinear model we have assumed that β_i is a $(k_i, 1)$ vector. In the present section we shall assume that all β vectors have the same order, i.e., $k_i = k$ $(i = 1, \dots, q)$. Examples of simple hypotheses are

$$H_0: \beta_3 = \beta_5 \quad \text{or} \quad H_0: \beta_2 = \beta_3 = \beta_5,$$

and of course

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_q.$$

An example of a composite hypothesis is

$$H_0: (\beta_1 = \beta_2 = \beta_3 \text{ and } \beta_4 = \beta_5).$$

Since we have discussed estimation of the constrained and unconstrained models (both linear and nonlinear) in sections 3 to 6, likelihood ratio tests can be straightforwardly performed. If, however, the model contains many parameters, execution of likelihood ratio tests can be quite costly. For that reason, let us take a closer look at the Wald test.

Define, as before,

$$\pi_1 = (1, 2, \dots, q_1) \quad \text{and} \quad \pi_2 = (q_1 + 1, q_1 + 2, \dots, q_1 + q_2),$$

with $q_1 + q_2 \leqq q$. Suppose we wish to test the hypothesis

$$H_0: \beta_i = \tilde{\beta}_h \quad \text{for} \quad i \in \pi_h, \; h = 1, 2, \tag{7.1}$$

employing a Wald statistic. That is, we wish to test the composite hypothesis that $\beta_1 = \beta_2 = \cdots = \beta_{q_1}$ and $\beta_{q_1+1} = \cdots = \beta_{q_1+2} = \cdots = \beta_{q_1+q_2}$. Define the kq_1 and kq_2 vectors

$$\bar{\beta}_1 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{q_1} \end{bmatrix} \quad \text{and} \quad \bar{\beta}_2 = \begin{bmatrix} \beta_{q_1+1} \\ \vdots \\ \beta_{q_1+q_2} \end{bmatrix}, \tag{7.2}$$

and for $h = 1, 2$ the $(q_h - 1, q_h)$ matrices

$$R_h = (s_{q_h - 1} \vdots -I_{q_h - 1}) = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & & \vdots \\ \vdots & \vdots & & \ddots & \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix}, \tag{7.3}$$

where s_{q_h-1} is a vector of (q_h-1) ones. Then (7.1) can be rewritten as

$$H_0: \left[\begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} \otimes I_k\right]\begin{pmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \end{pmatrix} = 0, \tag{7.4}$$

or even shorter as

$$H_0: \bar{R}\bar{\beta} = 0, \tag{7.5}$$

where

$$\bar{R} = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} \otimes I_k \quad \text{and} \quad \bar{\beta} = \begin{pmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \end{pmatrix}. \tag{7.6}$$

Let $\bar{\Omega}$ be the asymptotic covariance matrix of $\hat{\bar{\beta}}$, the ML estimator of $\bar{\beta}$. The appropriate Wald statistic to test H_0 is

$$\hat{\bar{\beta}}'\bar{R}'(\bar{R}\bar{\Omega}\bar{R}')^{-1}\bar{R}\hat{\bar{\beta}}, \tag{7.7}$$

which is asymptotically χ^2 distributed with $k[q_1+q_2-2]$ degrees of freedom. In the nonlinear case, (7.7) can — in general — not be simplified. The *linear* model (6.2), however, yields by Theorem 6.1

$$\bar{\Omega} = (1/q)\, s_{q_1+q_2} s'_{q_1+q_2} \otimes (A^{-1} - B^{-1}) + I_{q_1+q_2} \otimes B^{-1}. \tag{7.8}$$

Since $R_h s_{q_h} = 0$, $h=1,2$, we find that

$$\bar{R}\bar{\Omega}\bar{R}' = \begin{pmatrix} R_1R_1' & 0 \\ 0 & R_2R_2' \end{pmatrix} \otimes B^{-1}, \tag{7.9}$$

and

$$\bar{R}'(\bar{R}\bar{\Omega}\bar{R}')^{-1}\bar{R} = \begin{pmatrix} R_1'(R_1R_1')^{-1}R_1 & 0 \\ 0 & R_2'(R_2R_2')^{-1}R_2 \end{pmatrix} \otimes B. \tag{7.10}$$

At this point we need the following lemma.

Lemma 7.1. Let A be a non-singular $(q-1, q-1)$ matrix and let a be a $(q-1, 1)$ vector. Define the $(q-1, q)$ matrix $B = (a \,\vdots\, -A)$. Then,

$$B'(BB')^{-1}B = I_q - (1/b'b)bb',$$

where

$$b = \begin{pmatrix} 1 \\ A^{-1}a \end{pmatrix}. \qquad \square$$

In the special case of Lemma 7.1 where

$$A = I_{q_h - 1} \quad \text{and} \quad a = s_{q_h - 1},$$

we have

$$B = R_h \quad \text{and} \quad b = s_{q_h}.$$

Hence,

$$R_h(R_h R_h')^{-1} R_h' = I_{q_h} - (1/q_h)\, s_{q_h} s_{q_h}' \equiv I_{q_h} - M_h.$$

The Wald statistic (7.7) now reads

$$\hat{\bar{\beta}}' \begin{pmatrix} (I - M_1) \otimes B & 0 \\ 0 & (I - M_2) \otimes B \end{pmatrix} \hat{\bar{\beta}}$$

$$= \hat{\bar{\beta}}_1' [(I - M_1) \otimes B] \hat{\bar{\beta}}_1 + \hat{\bar{\beta}}_2' [(I - M_2) \otimes B] \hat{\bar{\beta}}_2 \equiv W_1 + W_2,$$

with

$$W_h = \sum_{j \in \pi_h} \left(\hat{\beta}_j' B \hat{\beta}_j \right) - \frac{1}{q_h} \left(\sum_{j \in \pi_h} \hat{\beta}_j \right)' B \left(\sum_{j \in \pi_h} \hat{\beta}_j \right), \qquad h = 1, 2.$$

Summarizing we have proved the following theorem.

Theorem 7.1. Consider the linear regression model (6.2) under the Assumption 3.1 (two error component matrices). Let $\hat{\beta}_1, \ldots, \hat{\beta}_q$ be the ML estimators for $\beta_1, \ldots, \beta_q$ without imposing any equality constraints. The Wald statistic to test the hypothesis

$$H_0 : \beta_i = \tilde{\beta}_h \quad \text{for} \quad i \in \pi_h \tag{7.11}$$

is

$$W_h \equiv \sum_{j \in \pi_h} (\hat{\beta}_j' B \hat{\beta}_j) - (1/q_h) \left(\sum_{j \in \pi_h} \hat{\beta}_j \right)' B \left(\sum_{j \in \pi_h} \hat{\beta}_j \right), \tag{7.12}$$

where

$$B = \sum_{t=1}^{T} X_t' \Delta^{-1} X_t.$$

W_h is — under H_0 — asymptotically χ^2 distributed with $k\,(q_h - 1)$ degrees of freedom. The Wald statistic to test the composite hypothesis

$$H_0 : \beta_i = \tilde{\beta}_h \quad \text{for} \quad i \in \pi_h, \qquad h \in H, \tag{7.13}$$

where H is a subset of $\{1,2,\ldots,r\}$, *is*

$$W_H \equiv \sum_{h\in H} W_h, \tag{7.14}$$

and is — under H_0 *— asymptotically* χ^2 *distributed with* $k\sum_{h\in H}(q_h-1)$ *degrees of freedom.* □

Thus, if the model under consideration is the linear regression model (6.2), testing of *all possible* equality restrictions on the β_i vectors is feasible, using the Wald statistic of Theorem 7.1. Execution of these tests is extremely inexpensive. Notice that Theorem 7.1 remains true without any alteration, when Assumption 3.1 (two error component matrices) is replaced by the simpler assumption that $u_{it}\simeq N_p(0,\Delta)$, i.e., the assumption that $\Gamma=0$ a priori.

8. An extension of the two components model

So far, we have discussed the nonlinear model (3.1)

$$y_{it}=f_i(X_{it},\beta_i)+u_{it}, \qquad i=1,\ldots,q, \quad t=1,\ldots,T,$$

and, as a special case, the linear model (6.2)

$$y_{it}=X_t\beta_i+u_{it}, \qquad i=1,\ldots,q, \quad t=1,\ldots,T.$$

In both cases, the error vectors u_{it} were supposed to decompose as $u_{it}=e_t+\varepsilon_{it}$ with e_t and ε_{it} independently distributed as $N_p(0,\Gamma)$ and $N_p(0,\Delta)$ respectively (Assumption 3.1).

This assumption on the error structure is (and obviously has to be) restrictive. The question arises whether it is *too* restrictive. In the present section we will discuss an extension of the error structure (3.1). This is laid down in:

Assumption 8.1. The error vectors u_{it} decompose as $u_{it}=e_t+\varepsilon_{it}$, where e_t and ε_{it} are independently distributed as

$$e_t\simeq N_p(0,\Gamma) \quad \text{and} \quad \varepsilon_{it}=N_p(0,\lambda_i\Delta),$$

with Γ positive semidefinite, Δ positive definite, both of order p, and λ_i $(i=1,\ldots,q)$ positive.

The only difference between Assumptions 3.1 and 8.1 is that the latter assumption allows the covariance matrix of ε_{it} to differ for each i (industry, say) by a factor of proportionality. The following rationale for introducing

Assumption 8.1 can be offered. Suppose we have observations on p cost shares during T years for each of q industries. The natural starting point is to analyze each industry separately by applying a Zellner-type estimation. In terms of our Assumptions 3.1 and 3.8, this means that $u_{it} \simeq N_p(0, \Delta_i)$. Assumption 3.8, combining the error component approach with the possibility that Δ differs from industry to industry (by a factor of proportionality), thus seems potentially useful.

It should be noted that, without a further constraint, only ratios of λ's can be identified, but not the λ's themselves. For that reason we normalize the λ's in the following way that will prove convenient.

Normalization 8.1. $\operatorname{tr} \Lambda^{-1} = q$, where Λ is the diagonal matrix with $\lambda_1, \dots, \lambda_q$ on the diagonal.

Proceeding now as in section 3 we find [eq. (3.5)][9]

$$\Omega = s_q s_q' \otimes \Gamma + \Lambda \otimes \Delta. \tag{8.1}$$

From Corollary 2.1 follows its determinant [eq. (3.8)]

$$|\Omega| = |\Lambda|^p |W| |\Delta|^{q-1}, \tag{8.2}$$

and its inverse [eq. (3.9)]

$$\Omega^{-1} = \Lambda^{-1} M \Lambda^{-1} \otimes (W^{-1} - \Delta^{-1}) + \Lambda^{-1} \otimes \Delta^{-1}. \tag{8.3}$$

Hence, Theorem 3.1 should be altered to:

Theorem 8.1 [3.1]. The loglikelihood associated with the nonlinear regression model (3.1) under the Assumption 8.1, is $\mathscr{L} = \sum_{t=1}^{T} \mathscr{L}_t$, *with*

$$\begin{aligned}\mathscr{L}_t = \gamma + \tfrac{1}{2} p \sum_{i=1}^{q} \log |\lambda_i^{-1}| - \tfrac{1}{2} \log |W| \\ - \tfrac{1}{2}(q-1) \log |\Delta| - \tfrac{1}{2} \operatorname{tr} V_t \Lambda^{-1} M \Lambda^{-1} V_t' W^{-1} \\ - \tfrac{1}{2} \operatorname{tr} V_t (\Lambda^{-1} - \Lambda^{-1} M \Lambda^{-1}) V_t' \Delta^{-1},\end{aligned} \tag{8.4}$$

[9]The notation indicates that formula (8.1) is the counterpart of formula (3.5) when Assumption 3.1 is replaced by Assumption 8.1. Hence, for $\Lambda = I_q$, (8.1) reduces to (3.5).

or equivalently,

$$\mathscr{L}_t=\gamma+\tfrac{1}{2}p\sum_{i=1}^{q}\log|\lambda_i^{-1}|-\tfrac{1}{2}\log|W|-\tfrac{1}{2}(q-1)\log|\Delta|$$

$$-\tfrac{1}{2}\sum_{i=1}^{q}\lambda_i^{-1}u_{it}'\Delta^{-1}u_{it}-(1/2q)\sum_{i,j=1}^{q}\lambda_i^{-1}\lambda_j^{-1}u_{it}'(W^{-1}-\Delta^{-1})u_{jt}. \qquad \square \quad (8.5)$$

Maximization of $\mathscr{L}$ is to be performed under the constraints $\lambda_i>0$ $(i=1,\ldots,q)$, $\sum_{i=1}^{q}\lambda_i^{-1}=q$, and Γ and Δ positive (semi)definite. The normalization $\sum_i\lambda_i^{-1}=q$ can be incorporated by substituting

$$\lambda_q^{-1}=q-\sum_{i=1}^{q-1}\lambda_i^{-1}. \tag{8.6}$$

To account for the inequality constraints $\lambda_i>0$ we define

$$\alpha_l=\sum_{i=1}^{q-1}|\lambda_i^{-1}|+\left|q-\sum_{i=1}^{q-1}\lambda_i^{-1}\right|-q. \tag{8.7}$$

Hence, $\alpha_l=0$ if $\lambda_i\geqq 0$ $(i=1,\ldots,q)$ and $\alpha_l<0$ otherwise. Let μ be some large positive number. Then maximization of

$$\mathscr{L}^*=\mathscr{L}-\mu\alpha_l, \tag{8.8}$$

will yield the required ML estimates. The same comments that followed Theorem 3.1 apply here, in particular as to the required positive (semi)definiteness of Γ and Δ.

The loglikelihood of the extended model (under Assumption 8.1) can be concentrated with respect to Γ and Δ, but *not*, it seems, with respect to Λ. This leads to Theorem 8.2 which is the counterpart of Theorem 4.1.

Theorem 8.2 [4.1]. Consider the model (3.1). Under Assumption 8.1 (instead of Assumption 3.1), Theorem 4.1 remains true if the following alterations are made

$$\phi(\beta_1,\ldots,\beta_q,\Lambda)=p\sum_{i=1}^{q}\log|\lambda_i|+\log\left|\sum_{t=1}^{T}V_t\Lambda^{-1}M\Lambda^{-1}V_t'\right|$$

$$+(q-1)\log\left|\sum_{t=1}^{T}V_t(\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})V_t'\right|, \tag{8.9}$$

$$\hat{\Delta}=\frac{1}{(q-1)T}\sum_{t=1}^{T}\hat{V}_t(\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})\hat{V}_t', \tag{8.10}$$

$$\hat{\Gamma}=\frac{1}{q(q-1)T}\sum_{t=1}^{T}\hat{V}_t(q\Lambda^{-1}M\Lambda^{-1}-\Lambda^{-1})\hat{V}_t', \tag{8.11}$$

$$\hat{W}=\frac{1}{T}\sum_{t=1}^{T}\hat{V}_t\Lambda^{-1}M\Lambda^{-1}\hat{V}_t', \tag{8.12}$$

$$\hat{\mathscr{L}}=\tfrac{1}{2}pqT\left[\log T+\frac{q-1}{q}\log(q-1)-\log 2\pi-1\right]$$

$$-\tfrac{1}{2}T\phi(\hat{\beta}_1,\dots,\hat{\beta}_q,\hat{\Lambda}).\quad\square \tag{8.13}$$

Notice that $\hat{\Gamma}$ from (8.11) — just as $\hat{\Gamma}$ from (4.9) — is not necessarily positive semidefinite. Turning now to the information matrix, we have the following adaption of Theorem 5.1.

Theorem 8.3 [5.1]. *Consider the model* (3.1). *Under Assumption 8.1 (instead of Assumption 3.1), Theorem 5.1 remains true if* C_{ij} *and* D_i *are replaced by* C_{ij}^* *and* D_i^*, *where*

$$C_{ij}^*=\lambda_i^{-1}\lambda_j^{-1}C_{ij}\quad\text{and}\quad D_i^*=\lambda_i^{-1}D_i.\quad\square$$

Theorem 8.3 gives the information matrix (and hence the 'asymptotic covariance matrix') for $\beta_1,\dots,\beta_q$. The natural way to proceed would be to derive the information matrix for Λ, Γ, and Δ. The relevant calculations, however, turn out to be extremely cumbersome. The counterpart of Theorem 5.2 is therefore not presented.

In the linear case $y_{it}=X_t\beta_i+u_{it}$, we can again solve the β's explicitly in terms of Λ, Γ, and Δ (Theorem 8.4) and perform the relevant Wald tests (Theorem 8.5).

Theorem 8.4 [6.1]. *Consider the linear regression model* (6.2). *Under the Assumptions 8.1 (instead of Assumption 3.1) and 6.1, Theorem 6.1 remains true if* q_h, $\bar{y}_t$, *and* $\bar{y}_{ht}$ *are replaced by* q_h^*, $\bar{y}_t^*$, *and* $\bar{y}_{ht}^*$, *where*

$$q_h^*=\sum_{j\in\pi_h}\lambda_j^{-1},\qquad \bar{y}_t^*=\frac{1}{q}\sum_{j=1}^{q}\lambda_j^{-1}y_{jt},\qquad \bar{y}_{ht}^*=\frac{1}{q_h^*}\sum_{j\in\pi_h}\lambda_j^{-1}y_{jt}.\quad\square$$

Theorem 8.5 [7.1]. *Consider the linear regression model* (6.2). *Under Assumption 8.1 (instead of Assumption 3.1), Theorem 7.1 remains true if* W_h *is* *replaced by*

$$W_h^* = \sum_{j\in\pi_h} \lambda_j^{-1}\hat\beta_j' B\hat\beta_j - (1/q_h^*)\left(\sum_{j\in\pi_h}\lambda_j^{-1}\hat\beta_j\right)' B\left(\sum_{j\in\pi_h}\lambda_j^{-1}\hat\beta_j\right), \tag{8.14}$$

where

$$q_h^* = \sum_{j\in\pi_h}\lambda_j^{-1}. \quad \square$$

To end our discussion of the extended two components model we have to formulate corollaries for the case where $\Gamma=0$. The error vectors u_{it} then follow the distribution

$$u_{it} \simeq N_p(0, \lambda_i \Delta),$$

with Δ positive definite and $\lambda_1,\dots,\lambda_q$ positive. This specification of the error structure, clearly, is a generalization of the standard Zellner-type of estimating seemingly unrelated regressions, and therefore interesting on its own account. The following four corollaries describe the estimation in this case without further discussion.

Corollary 8.1 [3.1]. *The loglikelihood associated with the nonlinear regression model* (3.1) *under the assumption that* $u_{it}\simeq N_p(0,\lambda_i\Delta)$, Δ *positive definite, is* $\mathscr{L}=\sum_{t=1}^T \mathscr{L}_t$ *with*

$$\mathscr{L}_t = \gamma - \tfrac{1}{2}p\sum_{i=1}^q \log|\lambda_i| - \tfrac{1}{2}q\log|\Delta| - \tfrac{1}{2}\operatorname{tr} V_t \Lambda^{-1} V_t' \Delta^{-1}$$

$$= \gamma - \tfrac{1}{2}p\sum_{i=1}^q \log|\lambda_i| - \tfrac{1}{2}q\log|\Delta| - \tfrac{1}{2}\sum_{i=1}^q \lambda_i^{-1} u_{it}'\Delta^{-1}u_{it}. \quad \square \tag{8.15}$$

Corollary 8.2 [4.1]. *Consider the model* (3.1). *Under the assumption that* $u_{it}\simeq N_p(0,\lambda_i\Delta)$, Δ *positive definite, Corollary 4.1 remains true if the following alterations are made*

$$\phi(\beta_1,\dots,\beta_q,\Lambda) = p\sum_{i=1}^q \log|\lambda_i| + \log\left|\sum_{t=1}^T V_t\Lambda^{-1}V_t'\right|,$$

$$\hat\Delta = (1/qT)\sum_{t=1}^T \hat V_t \Lambda^{-1}\hat V_t'$$

$$\hat{\mathscr{L}} = \tfrac{1}{2}pqT[\log qT - \log 2\pi - 1] - \tfrac{1}{2}qT\phi(\hat\beta_1,\dots,\hat\beta_q,\hat\Lambda). \quad \square$$

Corollary 8.3 [*5.1*]. *Consider the model* (*3.1*). *Under the assumption that* $u_{it} \simeq N_p(0, \lambda_i \Delta)$, Δ *positive definite, the information matrix for* β *is*

$$\Psi_T^{(\beta)} = \begin{bmatrix} \lambda_1^{-1} D_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_q^{-1} D_q \end{bmatrix}, \tag{8.16}$$

where D_i $(i=1,\ldots,q)$ *is defined in* (*5.4*). *Hence*,

$$\text{as.var}(\hat{\beta}_i) = \lambda_i D_i^{-1} \quad \textit{and} \quad \text{as.cov}(\hat{\beta}_i; \hat{\beta}_j) = 0. \quad \square$$

Corollary 8.4 [*6.1*]. *Consider the linear regression model* (*6.2*). *Under the assumption that* $u_{it} \simeq N(0, \lambda_i \Delta)$, Δ *positive definite, and Assumption 6.1, Corollary 6.1 remains true if* q_h *and* $\bar{y}_{ht}$ *are replaced by* q_h^* *and* $\bar{y}_{ht}^*$, *where*

$$q_h^* = \sum_{j \in \pi_h} \lambda_j^{-1} \quad \textit{and} \quad \bar{y}_{ht}^* = (1/q_h^*) \sum_{j \in \pi_h} y_{jt}. \quad \square$$

Appendix 1: The duplication matrix, traces, and differentials

Differentiation with respect to symmetric matrices is a delicate matter. The crux lies in the fact that we should only differentiate with respect to the *distinct* elements of a symmetric matrix. Let A be a (p,p) matrix, then $v(A)$ will denote the $\frac{1}{2}p(p+1)$-vector that is obtained from vec A by eliminating all supradiagonal elements of A. For example, if $p=3$,

$$\text{vec}\, A = (a_{11}\, a_{21}\, a_{31}\, a_{12}\, a_{22}\, a_{32}\, a_{13}\, a_{23}\, a_{33})',$$

and

$$v(A) = (a_{11}\, a_{21}\, a_{31}\, a_{22}\, a_{32}\, a_{33})'.$$

Thus, $v(A)$ contains the *distinct* elements of a symmetric (p,p) matrix. The $(p^2, \frac{1}{2}p(p+1))$ *duplication matrix* D performs for *symmetric* A the transformation

$$Dv(A) = \text{vec}\, A. \tag{A.1}$$

The duplication matrix is studied extensively in Henderson and Searle (1979) and Magnus and Neudecker (1980). D is uniquely determined by (A.1), and has full column rank $\frac{1}{2}p(p+1)$. Hence, $D'D$ is non-singular and D^+, the Moore–Penrose inverse of D, equals

$$D^+ = (D'D)^{-1} D.$$

For $p=3$, we have

$$D=\left[\begin{array}{ccc|cc|c} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \hline \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \hline \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{array}\right], \qquad D^{+\prime}=\left[\begin{array}{ccc|cc|c} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \frac{1}{2} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \frac{1}{2} & \cdot & \cdot & \cdot \\ \hline \cdot & \frac{1}{2} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \frac{1}{2} & \cdot \\ \hline \cdot & \cdot & \frac{1}{2} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \frac{1}{2} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{array}\right].$$

For a (p,p) matrix A, we define the $(\frac{1}{2}p(p+1), \frac{1}{2}p(p+1))$ matrices

$$P(A)=D'(A\otimes A)\,D \quad \text{and} \quad R(A)=D^{+}(A\otimes A)D^{+\prime}.$$

From Magnus and Neudecker (1980) we then have for non-singular A,

$$P(A^{-1})\,R(A)=I, \tag{A.2}$$

and

$$|R(A)|=2^{-\frac{1}{2}p(p-1)}|A|^{p+1}. \tag{A.3}$$

The following two results on traces are used:

$$(\operatorname{vec} A)'\operatorname{vec} B=\operatorname{tr} A'B, \tag{A.4}$$

if A and B have the same order, and

$$(\operatorname{vec} A)'(B\otimes C)\operatorname{vec} D=\operatorname{tr} DB'A'C, \tag{A.5}$$

if the R.H.S. exists.

Finally, we apply the following standard facts in matrix differentiation [see Neudecker (1969)]:

For every matrix X and Y of appropriate orders,

$$\mathrm{d}(XY)=(\mathrm{d}X)Y+X\,\mathrm{d}Y, \qquad \mathrm{d}\operatorname{tr} XY=\operatorname{tr}(\mathrm{d}X)Y+\operatorname{tr} X\,\mathrm{d}Y.$$

For every non-singular X,

$$\mathrm{d}X^{-1}=-X^{-1}(\mathrm{d}X)X^{-1},$$

and if $|X|>0$,

$$\mathrm{d}\log|X|=\operatorname{tr} X^{-1}\,\mathrm{d}X.$$

Appendix 2: Proofs

Proof of Lemma 2.1. (i) is a well-known result [see e.g. Rao and Mitra (1971, lemma 5.4.1)]. (iv) follows from (i), and (iii) from (ii). Let us proof (ii). By Schur's Theorem [see Bellman (1970, p. 202)] there exist non-singular matrices S_i that reduce A_i to triangular form: $S_i^{-1}A_iS_i=P_i$ $(i=1,\dots,s)$ where P_i is a lower triangular (p,p) matrix containing the eigenvalues of A_i on its diagonal. The matrices M_i are symmetric and idempotent. Hence, there exist (n,r_i) matrices T_i such that

$$M_i=T_iT_i', \qquad T_i'T_i=I_{r_i}, \qquad i=1,\dots,s.$$

The r_i columns of T_i are of course the eigenvectors of M_i associated with its unit eigenvalues. From $M_iM_j=0$ $(i\neq j)$ follows

$$T_i'T_j=0, \qquad i\neq j.$$

Let the (r_i,n) matrices Z_i be defined as

$$Z_i=[0 \,\vdots\, I_{r_i} \,\vdots\, 0], \qquad i=1,\dots,s,$$

where the first null-matrix has $(r_1+r_2+\cdots+r_{i-1})$ columns, and the second $(r_{i+1}+r_{i+2}+\cdots+r_s)$. Then,

$$\sum_{i=1}^{s} Z_i'Z_i=I_n.$$

Using these results, it is now easy to verify that the (np,np) matrices

$$Q_1=\sum_{i=1}^{s}(T_iZ_i\otimes S_i) \quad\text{and}\quad Q_2=\sum_{i=1}^{s}(S_i\otimes T_iZ_i)$$

are non-singular with inverses

$$Q_1^{-1}=\sum_{i=1}^{s}(Z_i'T_i'\otimes S_i^{-1}) \quad\text{and}\quad Q_2^{-1}=\sum_{i=1}^{s}(S_i^{-1}\otimes Z_i'T_i'),$$

and reduce Ω_1 and Ω_2 to triangular form:

$$Q_1^{-1}\Omega_1Q_1=\sum_{i=1}^{s}(Z_i'Z_i\otimes P_i),$$

$$Q_2^{-1}\Omega_2Q_2=\sum_{i=1}^{s}(P_i\otimes Z_i'Z_i).$$

The result follows. □

Proof of Lemma 2.2. First assume that B is non-singular. Then

$$G=(L\otimes B)\,[I\otimes I+L^{-1}ab'\otimes B^{-1}A].$$

Let λ_i $(i=1,\dots,p)$ be the eigenvalues of $B^{-1}A$. Then $L^{-1}ab'\otimes B^{-1}A$ has eigenvalues $\alpha\lambda_i$ $(i=1,\dots,p)$ and $p(q-1)$ zeros. Hence, $I\otimes I+L^{-1}ab'\otimes B^{-1}A$ has eigenvalues $1+\alpha\lambda_i$ $(i=1,\dots,p)$ and 1, so that

$$|I\otimes I+L^{-1}ab'\otimes B^{-1}A|=\prod_{i=1}^{p}(1+\alpha\lambda_i)=|I+\alpha B^{-1}A|.$$

Therefore,

$$|G|=|L|^p|B|^q|I+\alpha B^{-1}A|=|L|^p|B|^{q-1}|C|.$$

In case B is singular, we obtain $|G|=0$ starting with $B+\varepsilon I$, where ε is small and $B+\varepsilon I$ is non-singular. This yields

$$|L\otimes(B+\varepsilon I)+ab'\otimes A|=|L|^p|B+\varepsilon I|^{q-1}|C+\varepsilon I|.$$

Letting $\varepsilon\to 0$ gives the desired result. If L, B and C are all non-singular, one easily verifies that $GG^{-1}=I\otimes I$. Further, $C^{-1}AB^{-1}=B^{-1}AC^{-1}$, since

$$CB^{-1}A=(I+\alpha AB^{-1})A=A(I+\alpha B^{-1}A)=AB^{-1}C.$$

Finally,

$$C^{-1}AB^{-1}=\frac{1}{\alpha}(B^{-1}-C^{-1})\quad\text{if}\quad\alpha\neq 0,$$

since

$$C(B^{-1}-C^{-1})=(B+\alpha A)B^{-1}-I=\alpha AB^{-1}.\quad\square$$

Proof of Corollary 2.1. Immediate from Lemma 2.2. $\square$

Proof of Theorem 3.1. We write (3.1) as

$$y_t=f(X_t,\beta)+u_t,\qquad t=1,\dots,T,\tag{A.6}$$

with

$$y_t=\begin{bmatrix}y_{1t}\\ y_{2t}\\ \vdots\\ y_{qt}\end{bmatrix},\qquad u_t=\begin{bmatrix}u_{1t}\\ u_{2t}\\ \vdots\\ u_{qt}\end{bmatrix},\qquad \beta=\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_q\end{bmatrix},$$

and

$$X_t=(X_{1t}X_{2t}\dots X_{qt}).$$

The $(pq, 1)$ vectors u_t are identically and independently distributed as $N(0, \Omega)$, where Ω is given in (3.7). Hence, the probability density of y_t takes the form

$$(2\pi)^{-pq/2}|\Omega|^{-\frac{1}{2}}\exp -\tfrac{1}{2}u_t'\Omega^{-1}u_t,$$

and the loglikelihood is

$$\begin{aligned}\mathscr{L}_t &= -\tfrac{1}{2}pq\log 2\pi - \tfrac{1}{2}\log|\Omega| - \tfrac{1}{2}u_t'\Omega^{-1}u_t = \gamma - \tfrac{1}{2}\log|W| \\ &\quad -\tfrac{1}{2}(q-1)\log|\Delta| - \tfrac{1}{2}u_t'(M\otimes W^{-1} + (I-M)\otimes\Delta^{-1})u_t,\end{aligned}$$

using (3.8) and (3.9). Now, by (3.3) and (A.5),

$$u_t'(M\otimes W^{-1})u_t = \operatorname{tr} V_t M V_t' W^{-1},$$

and

$$u_t'((I-M)\otimes\Delta^{-1})u_t = \operatorname{tr} V_t(I-M)V_t'\Delta^{-1},$$

which proves (3.11). Also, one verifies that

$$u_t'(I\otimes\Delta^{-1})u_t = \sum_{i=1}^{q} u_{it}'\Delta^{-1}u_{it},$$

and

$$u_t'(M\otimes(W^{-1}-\Delta^{-1}))u_t = (1/q)\sum_{i,j=1}^{q} u_{it}'(W^{-1}-\Delta^{-1})u_{jt}.$$

This finishes the proof. □

Proof of Corollary 3.1. For $\Gamma=0$, we have $W=\Delta$, and Theorem 3.1 reduces to Corollary 3.1. □

Proof of Theorem 4.1. See text. □

Proof of Corollary 4.1. The proof is similar to the proof of Theorem 4.1, but simpler. Starting from (3.22),

$$\mathscr{L}_t = \gamma - \tfrac{1}{2}q\log|\Delta| - \tfrac{1}{2}\operatorname{tr} V_t V_t'\Delta^{-1}.$$

Totally differentiating with respect to Δ leads to

$$\mathrm{d}\mathscr{L}_t = \tfrac{1}{2}\operatorname{tr}\Delta^{-1}(V_tV_t' - q\Delta)\Delta^{-1}\,\mathrm{d}\Delta,$$

and hence

$$\mathrm{d}\mathscr{L}=\sum_{t=1}^{T}\mathrm{d}\mathscr{L}_t=\tfrac{1}{2}\operatorname{tr}\Delta^{-1}\left(\sum_{t=1}^{T}V_tV_t'-qT\Delta\right)\Delta^{-1}\mathrm{d}\Delta.$$

We find the ML equation for Δ to be

$$\Delta=(1/qT)\sum_{t=1}^{T}V_tV_t'.$$

Concentrating with respect to Δ leads to this well-known corollary. □

Proof of Theorem 5.1. Model (3.1),

$$y_{it}=f_i(X_{it},\beta_i)+u_{it},\qquad i=1,\ldots,q,\quad t=1,\ldots,T,$$

can be written as in (A.6)

$$y_t=f(X_t,\beta)+u_t,\qquad t=1,\ldots,T.$$

and even more compactly as

$$y=g(X,\beta)+u, \tag{A.7}$$

where

$$y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_T\end{bmatrix},\qquad g(X,\beta)=\begin{bmatrix}f(X_1,\beta)\\f(X_2,\beta)\\\vdots\\f(X_T,\beta)\end{bmatrix},\qquad u=\begin{bmatrix}u_1\\u_2\\\vdots\\u_T\end{bmatrix}.$$

From Heymans and Magnus (1979), we know that the information matrix for β is the $(\sum_{i=1}^{q}k_i,\sum_{i=1}^{q}k_i)$ matrix

$$\Psi_T^{(\beta)}=H'\Omega^{*-1}H,$$

where Ω^* is the (pqT,pqT) covariance matrix of u, and H is the $(pqT,\sum_{i=1}^{q}k_i)$ matrix

$$H=\left(\frac{\partial g(X,\beta)}{\partial\beta}\right)'.$$

We first evaluate H. We have

$$H'=\begin{bmatrix} \dfrac{\partial f_1(X_{11},\beta_1)}{\partial\beta_1} & \cdots & \dfrac{\partial f_q(X_{q1},\beta_q)}{\partial\beta_1} & \cdots & \dfrac{\partial f_1(X_{1T},\beta_1)}{\partial\beta_1} & \cdots & \dfrac{\partial f_q(X_{qT},\beta_q)}{\partial\beta_1} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \dfrac{\partial f_1(X_{11},\beta_1)}{\partial\beta_q} & \cdots & \dfrac{\partial f_q(X_{q1},\beta_q)}{\partial\beta_q} & \cdots & \dfrac{\partial f_1(X_{1T},\beta_1)}{\partial\beta_q} & \cdots & \dfrac{\partial f_q(X_{qT},\beta_q)}{\partial\beta_q} \end{bmatrix}$$

$$=\left[\begin{array}{ccc:c:ccc} H'_{11} & & 0 & & H'_{1T} & & 0 \\ & \ddots & & \cdots & & \ddots & \\ 0 & & H'_{q1} & & 0 & & H'_{qT} \end{array}\right],$$

where H_{it} is defined in (5.2). Let e_t be the tth column of I_T, and E_{ij} the (q,q) matrix with unity in its ijth position and zeroes elsewhere. Then,

$$H'=\sum_{i=1}^{q}\sum_{t=1}^{T}(e_t'\otimes E_{ii}\otimes H_{it}'). \tag{A.8}$$

The covariance matrix Ω^* equals

$$\Omega^*=I_T\otimes\Omega,$$

where Ω is defined in (3.7). Hence

$$\Omega^{*-1}=I_T\otimes M\otimes(W^{-1}-\Delta^{-1})+I_T\otimes I_q\otimes\Delta^{-1}, \tag{A.9}$$

and

$$\begin{aligned}\Psi_T^{(\beta)} &= H'\Omega^{*-1}H \\ &= \sum_{i,j=1}^{q}\sum_{t,s=1}^{T}(e_t'\otimes E_{ii}\otimes H_{it}') \\ &\quad\times[I_T\otimes M\otimes(W^{-1}-\Delta^{-1})+I_T\otimes I_q\otimes\Delta^{-1}](e_s\otimes E_{jj}\otimes H_{js}) \\ &= \sum_{i,j=1}^{q}\sum_{t,s=1}^{T} e_t'e_s\otimes E_{ii}ME_{jj}\otimes H_{it}'(W^{-1}-\Delta^{-1})H_{js} \\ &\quad+\sum_{i,j=1}^{q}\sum_{t,s=1}^{T} e_t'e_s\otimes E_{ii}E_{jj}\otimes H_{it}'\Delta^{-1}H_{js}\end{aligned}$$

$$= \sum_{i,j=1}^{q} \sum_{t=1}^{T} (1/q)\, E_{ij} \otimes H'_{it} (W^{-1} - \Delta^{-1}) H_{jt}$$

$$+ \sum_{i=1}^{q} \sum_{t=1}^{T} E_{ii} \otimes H'_{it} \Delta^{-1} H_{it}$$

$$= (1/q) \sum_{i,j=1}^{q} E_{ij} \otimes \left(\sum_{t=1}^{T} H'_{it} (W^{-1} - \Delta^{-1}) H_{jt} \right)$$

$$+ \sum_{i=1}^{q} E_{ii} \otimes \left(\sum_{t=1}^{T} H'_{it} \Delta^{-1} H_{it} \right). \quad \square$$

Proof of Theorem 5.2. Recall from (4.5) the first differential of $\mathscr{L}$,

$$\mathrm{d}\mathscr{L} = (q/2)(\mathrm{d}v(\Gamma))' D' \operatorname{vec} \tilde{W} + \tfrac{1}{2} (\mathrm{d}v(\Delta))' D' \operatorname{vec} (\tilde{W} + \tilde{\Delta}).$$

We first calculate the expectations of the differentials of $\tilde{W}$ and $\tilde{\Delta}$,

$$\mathrm{d}\tilde{W} = (\mathrm{d}W^{-1}) \left(\sum_t V_t M V'_t - TW \right) W^{-1}$$

$$+ W^{-1} \left(\sum_t V_t M V'_t - TW \right) (\mathrm{d}W^{-1}) - TW^{-1} (\mathrm{d}W) W^{-1}.$$

Since $E V_t M V'_t = W$ [see (3.10)], we have

$$-\mathrm{E}\,\mathrm{d}\tilde{W} = TW^{-1} (\mathrm{d}W) W^{-1}.$$

Similarly,

$$-\mathrm{E}\,\mathrm{d}\tilde{\Delta} = (q-1) T \Delta^{-1} (\mathrm{d}\Delta) \Delta^{-1}.$$

The second differential of $\mathscr{L}$ is

$$\mathrm{d}^2 \mathscr{L} = (q/2)(\mathrm{d}v(\Gamma))' D' \operatorname{vec} \mathrm{d}\, \tilde{W} + \tfrac{1}{2} (\mathrm{d}v(\Delta))' D' (\operatorname{vec} \mathrm{d}\, \tilde{W} + \operatorname{vec} \mathrm{d}\, \tilde{\Delta}).$$

Hence

$$-\mathrm{E}\mathrm{d}^2\mathscr{L}=(qT/2)(\mathrm{d}v(\Gamma))'D'\operatorname{vec}[W^{-1}(\mathrm{d}W)W^{-1}]$$
$$+(T/2)(\mathrm{d}v(\Delta))'D'\operatorname{vec}[W^{-1}(\mathrm{d}W)W^{-1}+(q-1)\Delta^{-1}(\mathrm{d}\Delta)\Delta^{-1}]$$
$$=(qT/2)(\mathrm{d}v(\Gamma))'D'(W^{-1}\otimes W^{-1})\mathrm{d}\operatorname{vec}W$$
$$+(T/2)(\mathrm{d}v(\Delta))'D'(W^{-1}\otimes W^{-1})\mathrm{d}\operatorname{vec}W$$
$$+((q-1)T/2)(\mathrm{d}v(\Delta))'D'(\Delta^{-1}\otimes\Delta^{-1})\mathrm{d}\operatorname{vec}\Delta.$$

Since

$$\mathrm{d}W=\mathrm{d}\Delta+q\mathrm{d}\Gamma,$$

and

$$\mathrm{d}\operatorname{vec}\Delta=D\mathrm{d}v(\Delta),\qquad \mathrm{d}\operatorname{vec}\Gamma=D\mathrm{d}v(\Gamma),$$

we have

$$-\mathrm{E}\mathrm{d}^2\mathscr{L}=(qT/2)(\mathrm{d}v(\Gamma))'D'(W^{-1}\otimes W^{-1})D\mathrm{d}v(\Delta)$$
$$+(q^2T/2)(\mathrm{d}v(\Gamma))'D'(W^{-1}\otimes W^{-1})D\mathrm{d}v(\Gamma)$$
$$+(T/2)(\mathrm{d}v(\Delta))'D'(W^{-1}\otimes W^{-1})D\mathrm{d}v(\Delta)$$
$$+(qT/2)(\mathrm{d}v(\Delta))'D'(W^{-1}\otimes W^{-1})D\mathrm{d}v(\Gamma)$$
$$+((q-1)T/2)(\mathrm{d}v(\Delta))'D'(\Delta^{-1}\otimes\Delta^{-1})D\mathrm{d}v(\Delta).$$

The information matrix (5.7) then follows. One verifies the expression for the asymptotic covariance matrix (5.8) by direct multiplication, since $P(A^{-1})R(A)=I$, see (A.2). □

Proof of Corollary 5.1. Immediate from the fact that $\Gamma=0$ implies that $W=\Delta$. □

Proof of Theorem 6.1. Our starting point is (6.7),

$$y_t=(N\otimes X_t)\tilde{\beta}+u_t,\qquad t=1,\ldots,T,$$

where N is the blockdiagonal (q,r) matrix defined in (6.6). Let

$$y=\begin{bmatrix}y_1\\ \vdots\\ y_T\end{bmatrix},\qquad u=\begin{bmatrix}u_1\\ \vdots\\ u_T\end{bmatrix},\qquad X=\begin{bmatrix}N\otimes X_1\\ \vdots\\ N\otimes X_T\end{bmatrix},$$

then

$$N's_q = Qs_r = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_r \end{bmatrix}, \tag{A.13}$$

and

$$N'MN = (1/q)\,Qs_rs_r'Q \quad \text{and} \quad N'N = Q.$$

Let A and B be as in Theorem 6.1, then

$$X'\Omega^{*-1}X = Qs_rs_r'Q \otimes (1/q)(A-B) + Q \otimes B, \tag{A.14}$$

and its inverse follows from Lemma 2.2,

$$(X'\Omega^{*-1}X)^{-1} = (1/q)s_rs_r' \otimes (A^{-1}-B^{-1}) + Q^{-1} \otimes B^{-1}. \tag{A.15}$$

This is the asymptotic covariance matrix of $\hat{\hat{\beta}}$. In order to evaluate $X'\Omega^{*-1}y$, it is convenient to write y as

$$y = \sum_{t=1}^{T}\sum_{j=1}^{q}(e_t^{(T)} \otimes e_j^{(q)} \otimes y_{it}), \tag{A.16}$$

where $e_t^{(T)}$ and $e_j^{(q)}$ are the tth and jth column of I_T and I_q respectively. Now,

$$\begin{aligned}
X'\Omega^{*-1}y &= \sum_{t=1}^{T}(e_t^{(T)\prime} \otimes N' \otimes X_t')(I_T \otimes \Omega^{-1})\left(\sum_{s=1}^{T}\sum_{j=1}^{q} e_s^{(T)} \otimes e_j^{(q)} \otimes y_{js}\right) \\
&= \sum_{t=1}^{T}\sum_{j=1}^{q}(N' \otimes X_t')\Omega^{-1}(e_j^{(q)} \otimes y_{jt}) \\
&= \sum_{tj}(N' \otimes X_t')(M \otimes (W^{-1}-\Delta^{-1}) + I_q \otimes \Delta^{-1})(e_j^{(q)} \otimes y_{jt}) \\
&= \sum_{j=1}^{q} N'Me_j^{(q)} \otimes \left(\sum_t X_t'(W^{-1}-\Delta^{-1})y_{jt}\right) \\
&\quad + \sum_{j=1}^{q} N'e_j^{(q)} \otimes \left(\sum_t X_t'\Delta^{-1}y_{jt}\right) \\
&= \sum_{j=1}^{q} N'Me_j^{(q)} \otimes \left(\sum_t X_t'(W^{-1}-\Delta^{-1})y_{jt}\right) \\
&\quad + \sum_{h=1}^{r}\sum_{j\in\pi_h} N'e_j^{(q)} \otimes \left(\sum_t X_t'\Delta^{-1}y_{jt}\right).
\end{aligned}$$

Since

$$N'Me_j^{(q)} = \frac{1}{q} Qs_r \quad \text{and} \quad N'e_j^{(q)} = e_h^{(r)} \quad \text{for} \quad j \in \pi_h,$$

where $e_h^{(r)}$ is the hth column of I_r, it follows that

$$X'\Omega^{*-1}y = Qs_r \otimes (a-b) + \sum_{h=1}^{r} e_h^{(r)} \otimes q_h b_h, \tag{A.17}$$

with a, b, and b_h as in Theorem 6.1. Realizing that

$$s_r' Q s_r = q, \quad (1/q)\sum_{h=1}^{r} q_h b_h = b, \quad Q^{-1} e_h^{(r)} = (1/q_h)\, e_h^{(r)},$$

we find that

$$\begin{aligned}\tilde{\beta} &= (X'\Omega^{*-1}X)^{-1} X'\Omega^{*-1}y \\ &= s_r \otimes (A^{-1}a - B^{-1}b) + \sum_{h=1}^{r} e_h^{(r)} \otimes B^{-1} b_h,\end{aligned}$$

the hth subvector of which is

$$\tilde{\beta}_h = A^{-1}a + B^{-1}(b_h - b). \quad \square$$

Proof of Corollary 6.1. If $\Gamma = 0$, then $W = \Delta$, $A = B$, and $a = b$. $\square$

Proof of Lemma 7.1. Let $c = A^{-1}a$ and $C = (c \,\vdots\, -I_{q-1})$. Then

$$B = AC \quad \text{and} \quad B'(BB')^{-1}B = C'(CC')^{-1}C.$$

Now

$$CC' = I_{q-1} + cc' \quad \text{and} \quad (CC')^{-1} = I_{q-1} - \frac{cc'}{1+c'c}.$$

Hence

$$C'(CC')^{-1}C = \begin{pmatrix} 1 & 0 \\ 0 & I_{q-1} \end{pmatrix} - \frac{1}{1+c'c}\begin{pmatrix} 1 & c' \\ c & cc' \end{pmatrix} = I_q - \frac{1}{b'b} bb'. \quad \square$$

Proof of Theorem 7.1. See text. $\square$

Proof of Theorem 8.1. As the proof of Theorem 3.1, using (8.2) and (8.3), and the results

$$u_t'(\Lambda^{-1}M\Lambda^{-1}\otimes W^{-1})u_t=\operatorname{tr} V_t\Lambda^{-1}M\Lambda^{-1}V_t'W^{-1},$$

$$u_t'((\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})\otimes\Delta^{-1})u_t=\operatorname{tr} V_t(\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})V_t'\Delta^{-1},$$

$$u_t'(\Lambda^{-1}\otimes\Delta^{-1})u_t=\sum_{i=1}^{q}\lambda_i^{-1}u_{it}'\Delta^{-1}u_{it},$$

$$u_t'(\Lambda^{-1}M\Lambda^{-1}\otimes(W^{-1}-\Delta^{-1}))u_t$$

$$=(1/q)\sum_{i,j=1}^{q}\lambda_i^{-1}\lambda_j^{-1}u_{it}'(W^{-1}-\Delta^{-1})u_{jt}. \quad \square$$

Proof of Theorem 8.2. Totally differentiating (8.4) with respect to Γ and Δ yields

$$\mathrm{d}\mathscr{L}_t=-\tfrac{1}{2}\mathrm{d}\log|W|-\tfrac{1}{2}(q-1)\mathrm{d}\log|\Delta|-\tfrac{1}{2}\operatorname{tr} V_t\Lambda^{-1}M\Lambda^{-1}V_t'(\mathrm{d}W^{-1})$$

$$-\tfrac{1}{2}\operatorname{tr} V_t(\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})V_t'(\mathrm{d}\Delta^{-1})$$

$$=\tfrac{1}{2}\operatorname{tr} W^{-1}[V_t\Lambda^{-1}M\Lambda^{-1}V_t'-W]W^{-1}\mathrm{d}W$$

$$+\tfrac{1}{2}\operatorname{tr}\Delta^{-1}[V_t(\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})V_t'-(q-1)\Delta]\Delta^{-1}\mathrm{d}\Delta.$$

Hence

$$\mathrm{d}\mathscr{L}=\sum_{t=1}^{T}\mathrm{d}\mathscr{L}_t=\tfrac{1}{2}\operatorname{tr}\tilde{\tilde{W}}\mathrm{d}W+\tfrac{1}{2}\operatorname{tr}\tilde{\tilde{\Delta}}\mathrm{d}\Delta,$$

with

$$\tilde{\tilde{W}}=W^{-1}\left(\sum_{t=1}^{T}V_t\Lambda^{-1}M\Lambda^{-1}V_t'-TW\right)W^{-1},$$

and

$$\tilde{\tilde{\Delta}}=\Delta^{-1}\left(\sum_{t=1}^{T}V_t(\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})V_t'-(q-1)T\Delta\right)\Delta^{-1}.$$

Precisely as in the passage from (4.4) to (4.6), we find the ML equations for W and Δ,

$$W=(1/T)\sum_{t=1}^{T}V_t\Lambda^{-1}M\Lambda^{-1}V_t'$$

and

$$\Delta=(1/(q-1)T)\sum_{t=1}^{T} V_t(\Lambda^{-1}-\Lambda^{-1}M\Lambda^{-1})V_t'.$$

Substituting these expressions for W and Δ into $\mathscr{L}$ gives the concentrated loglikelihood. The results follow. □

Proof of Theorem 8.3. The proof is similar to the proof of Theorem 5.1. In particular, H' is as in (A.8),

$$H'=\sum_{i=1}^{q}\sum_{t=1}^{T}(e_t'\otimes E_{ii}\otimes H_{it}'),$$

and Ω^{*-1} is

$$\Omega^{*-1}=I_T\otimes\Lambda^{-1}M\Lambda^{-1}\otimes(W^{-1}-\Delta^{-1})+I_T\otimes\Lambda^{-1}\otimes\Delta^{-1}.$$

Hence

$$\begin{aligned}\Psi_T^{(\beta)}&=H'\Omega^{*-1}H\\&=(1/q)\sum_{i,j=1}^{q}E_{ij}\otimes\lambda_i^{-1}\lambda_j^{-1}\left(\sum_t H_{it}'(W^{-1}-\Delta^{-1})H_{jt}\right)\\&\quad+\sum_{i=1}^{q}E_{ii}\otimes\lambda_i^{-1}\left(\sum_t H_{it}'\Delta^{-1}H_{it}\right).\end{aligned}$$ □

Proof of Theorem 8.4. As the proof of Theorem 6.1, but now with

$$\Omega^{*-1}=I_T\otimes\Lambda^{-1}M\Lambda^{-1}\otimes(W^{-1}-\Delta^{-1})+I_T\otimes\Lambda^{-1}\otimes\Delta^{-1}.$$

We have

$$\begin{aligned}X'\Omega^{*-1}X&=N'\Lambda^{-1}M\Lambda^{-1}N\otimes\sum_t X_t'(W^{-1}-\Delta^{-1})X_t\\&\quad+N'\Lambda^{-1}N\otimes\sum_t X_t'\Delta^{-1}X_t.\end{aligned}$$

Let

$$Q^*=\begin{bmatrix}q_1^* & & 0\\ & \ddots & \\ 0 & & q_r^*\end{bmatrix}\quad\text{with}\quad q_h^*=\sum_{j\in\pi_h}\lambda_j^{-1},\quad h=1,\ldots,r.$$

Then

$$N'\Lambda^{-1}s_q=Q^*s_r,$$

$$N'\Lambda^{-1}M\Lambda^{-1}N=(1/q)\,Q^*s_rs_r'Q^*,$$

$$N'\Lambda^{-1}N=Q^*.$$

Hence

$$X'\Omega^{*-1}X=Q^*s_rs_r'\,Q^*\otimes(1/q)(A-B)+Q^*\otimes B,$$

and

$$(X'\Omega^{*-1}X)^{-1}=(1/q)s_rs_r'\otimes(A^{-1}-B^{-1})+Q^{*-1}\otimes B^{-1}.$$

Further

$$X'\Omega^{*-1}y=\sum_{j=1}^{q}N'\Lambda^{-1}\,M\Lambda^{-1}\,e_j^{(q)}\otimes\left(\sum_t X_t'(W^{-1}-\Delta^{-1})y_{jt}\right)$$

$$+\sum_{h=1}^{r}\sum_{j\in\pi_h}N'\Lambda^{-1}e_j^{(q)}\otimes\left(\sum_t X_t'\Delta^{-1}y_{jt}\right).$$

Since

$$N'\Lambda^{-1}M\Lambda^{-1}e_j^{(q)}=(1/q)\,\lambda_j^{-1}\,Q^*s_r$$

and

$$N'\Lambda^{-1}e_j^{(q)}=\lambda_j^{-1}\,e_h^{(r)}\quad\text{for}\quad j\in\pi_h,$$

it follows that

$$X'\Omega^{*-1}y=Q^*s_r\otimes\left(\sum_t X_t'(W^{-1}-\Delta^{-1})\bar{y}_t^*\right)$$

$$+\sum_{h=1}^{r}e_h^{(r)}\otimes q_h^*\left(\sum_t X_t'\Delta^{-1}\bar{y}_{ht}^*\right)$$

$$=Q^*s_r\otimes(a-b)+\sum_{h=1}^{r}e_h^{(r)}\otimes q_h^*\,b_h.$$

In the present case,

$$s_r'Q^*s_r=q,\quad(1/q)\sum_{h=1}^{r}q_h^*\,b_h=b,\quad Q^{*-1}e_h^{(r)}=(1/q_h^*)e_h^{(r)},$$

and the results follow. □

Proof of Theorem 8.5. The argument is as in section 7. The appropriate Wald statistic to test H_0 is (7.7),

$$\hat{\bar{\beta}}' \bar{R}'(\bar{R}\bar{\Omega}\bar{R}')^{-1} \bar{R} \hat{\bar{\beta}}.$$

By Theorem 8.4, and the definition of Q^*,

$$\bar{\Omega} = (1/q) s_{q_1+q_2} s'_{q_1+q_2} \otimes (A^{-1} - B^{-1}) + \bar{\Lambda} \otimes B^{-1},$$

where

$$\bar{\Lambda} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \quad \Lambda_1 = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{q_1} \end{bmatrix}, \quad \Lambda_2 = \begin{bmatrix} \lambda_{q_1+1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{q_1+q_2} \end{bmatrix}.$$

Since $R_h s_{q_h} = 0$ $(h = 1, 2)$, we find that

$$\bar{R}\bar{\Omega}\bar{R}' = \begin{bmatrix} R_1 \Lambda_1 R'_1 & 0 \\ 0 & R_2 \Lambda_2 R'_2 \end{bmatrix} \otimes B^{-1},$$

and

$$\bar{R}'(\bar{R}\bar{\Omega}\bar{R}')^{-1}\bar{R} = \begin{pmatrix} R'_1 (R_1 \Lambda_1 R'_1)^{-1} R_1 & 0 \\ 0 & R'_2 (R_2 \Lambda_2 R'_2)^{-1} R_2 \end{pmatrix} \otimes B.$$

Now, for $h = 1, 2$, let Λ_h be partitioned as

$$\Lambda_h = \begin{pmatrix} \gamma_h & 0 \\ 0 & C_h \end{pmatrix}.$$

Define

$$B_h \equiv R_h \Lambda_h^{\frac{1}{2}} = (\gamma_h^{\frac{1}{2}} s_{q_h - 1} \vdots -C_h^{\frac{1}{2}}).$$

Then, using Lemma 7.1,

$$\begin{aligned} R'_h (R_h \Lambda_h R'_h)^{-1} R_h &= \Lambda_h^{-\frac{1}{2}} B_h (B_h B'_h)^{-1} B'_h \Lambda_h^{-\frac{1}{2}} \\ &= \Lambda_h^{-\frac{1}{2}} [I_{q_h} - (1/q_h^*) \Lambda_h^{-\frac{1}{2}} s_{q_h} s'_{q_h} \Lambda_h^{-\frac{1}{2}}] \Lambda_h^{-\frac{1}{2}} \\ &= \Lambda_h^{-1} - (1/q_h^*) \Lambda_h^{-1} s_{q_h} s'_{q_h} \Lambda_h^{-1}. \end{aligned}$$

The Wald statistic (7.7) is $W_1^* + W_2^*$ with

$$W_h^* = \hat{\bar{\beta}}_h'[(\Lambda_h^{-1} - (1/q_h^*)\Lambda_h^{-1} s_{q_h} s_{q_h}' \Lambda_h^{-1}) \otimes B]\hat{\bar{\beta}}_h$$

$$= \sum_{j \in \pi_h} \lambda_j^{-1} \hat{\beta}_j' B \hat{\beta}_j - (1/q_h^*) \left(\sum_{j \in \pi_h} \lambda_j^{-1} \hat{\beta}_j \right)' B \left(\sum_{j \in \pi_h} \lambda_j^{-1} \hat{\beta}_j \right). \quad \square$$

Proof of Corollaries 8.1–8.4. The proofs are all simple and similar to their counterparts 3.1, 4.1, 5.1, and 6.1. $\square$

References

Amemiya, T., 1971, The estimation of the variances in a variance-components model, International Economic Review 12, 1–13.

Avery, R.B., 1977, Error components and seemingly unrelated regressions, Econometrica 45, 199–209.

Balestra, P., 1973, Best quadratic unbiased estimators of the variance-covariance matrix in normal regression, Journal of Econometrics 1, 17–28.

Balestra, P. and M. Nerlove, 1966, Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas, Econometrica 34, 585–612.

Baltagi, B.H., 1980, On seemingly unrelated regressions with error components, Econometrica 48, 1547–1551.

Bellman, R. 1970, Introduction to matrix analysis, 2nd ed. (McGraw-Hill, New York).

Chamberlain, G. and Z. Griliches, 1975, Unobservables with a variance-components structure: Ability, schooling and the economic success of brothers, International Economic Review 16, 422–450.

Darrough, M., R.A. Pollak and T.J. Wales, 1980, Taste change and stochastic structure: An analysis of three time series of household budget studies, Discussion paper no. 80-38 (University of British Columbia, Vancouver).

Fuller, W.A. and G.E. Battese, 1974, Estimation of linear functions with crossed-error structure, Journal of Econometrics 2, 67–78.

Henderson, H.V. and S.R. Searle, 1979, Vec and Vech operators for matrices, with some uses in Jacobians and multivariate statistics, Canadian Journal of Statistics 7, 65–81.

Heymans, R.D.H. and J.R. Magnus, 1979, Asymptotic properties of maximum likelihood estimators in a nonlinear model with unknown parameters in the disturbance covariance matrix, Paper presented at the European Meeting of the Econometric Society in Athens.

Kanemoto, Y., 1980, Price–quantity dynamics in a monopolistically competitive economy with small inventory costs, Discussion paper no. 83 (80–21) (University of Tsukuba, Japan).

Maddala, G.S., 1971, The use of variance components models in pooling cross-section and time-series data, Econometrica 39, 341–358.

Maddala, G.S. and T.D. Mount, 1973, A comparative study of alternative estimators for variance components models used in econometric applications, Journal of the American Statistical Association 68, 324–328.

Magnus, J.R., 1978, Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix, Journal of Econometrics 7, 281–312. Corrigenda, Journal of Econometrics 10, 261.

Magnus, J.R., 1980, Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood, Discussion paper no. 80–34 (University of British Columbia, Vancouver).